

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 16:24:42 ; Search time 1810.21 Seconds  
(without alignments)  
16077.036 Million cell updates/sec

Title: US-09-708-724A-3\_COPY\_1\_1000  
Perfect score: 1000  
Sequence: 1 agccagactaggagtgcgc.....cacacatagagtcacagagga 1000

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 2000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl:\*\*

- 1: gb\_ba:\*\*
- 2: gb\_htg:\*\*
- 3: gb\_in:\*\*
- 4: gb\_ov:\*\*
- 5: gb\_pat:\*\*
- 6: gb\_ph:\*\*
- 7: gb\_pl:\*\*
- 8: gb\_pr:\*\*
- 9: gb\_ro:\*\*
- 10: gb\_sts:\*\*
- 11: gb\_sy:\*\*
- 12: gb\_un:\*\*
- 13: gb\_vl:\*\*
- 14: gb\_vr:\*\*
- 15: em\_ba:\*\*
- 16: em\_fun:\*\*
- 17: em\_hum:\*\*
- 18: em\_in:\*\*
- 19: em\_mu:\*\*
- 20: em\_or:\*\*
- 21: em\_ov:\*\*
- 22: em\_pat:\*\*
- 23: em\_ph:\*\*
- 24: em\_pl:\*\*
- 25: em\_ro:\*\*
- 26: em\_sts:\*\*
- 27: em\_un:\*\*
- 28: em\_vl:\*\*
- 29: em\_vr:\*\*
- 30: em\_htg\_hum:\*\*
- 31: em\_htg\_inv:\*\*
- 32: em\_htg\_other:\*\*
- 33: em\_htg\_mus:\*\*
- 34: em\_htg\_pin:\*\*
- 35: em\_htg\_rod:\*\*
- 36: em\_htg\_mam:\*\*
- 37: em\_htg\_vrt:\*\*
- 38: em\_sy:\*\*
- 39: em\_htgo\_hum:\*\*
- 40: em\_htgo\_mus:\*\*
- 41: em\_htgo\_other:\*\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match % | Length | DB | ID       | Description        |
|------------|-------|---------------|--------|----|----------|--------------------|
| 1          | 1000  | 100.0         | 151049 | 9  | AC018558 | Homo sapi          |
| 2          | 1000  | 100.0         | 191768 | 9  | AC092357 | Homo sapi          |
| 3          | 306.4 | 30.6          | 167133 | 9  | AC026186 | Homo sapi          |
| c 4        | 306.4 | 30.6          | 201061 | 9  | AC007353 | Homo sapi          |
| c 5        | 279.8 | 28.0          | 144577 | 9  | AC023824 | Homo sapi          |
| c 6        | 269.4 | 26.9          | 70048  | 9  | HSU91320 | Human chrom        |
| c 7        | 269.4 | 26.9          | 129290 | 2  | AC020761 | Homo sapi          |
| c 8        | 269.4 | 26.9          | 167553 | 2  | AC114938 | Homo sapi          |
| c 9        | 269.4 | 26.9          | 185994 | 2  | AC002042 | Homo sapi          |
| c 10       | 269.4 | 26.9          | 214025 | 9  | AC007882 | Homo sapi          |
| c 11       | 269.4 | 26.9          | 227856 | 2  | AC007908 | Homo sapi          |
| c 12       | 232.2 | 23.2          | 129290 | 2  | AC020761 | Homo sapi          |
| c 13       | 121   | 12.1          | 57476  | 2  | AC107630 | Homo sapi          |
| c 14       | 89    | 8.9           | 176547 | 9  | AC092447 | Homo sapi          |
| c 15       | 87.4  | 8.7           | 194142 | 9  | AC092634 | Homo sapi          |
| c 16       | 85.2  | 8.5           | 44013  | 2  | AC087356 | Homo sapi          |
| c 17       | 79.8  | 8.0           | 104660 | 9  | AP000347 | AP000347 Pan trogl |
| c 18       | 79.8  | 8.0           | 208825 | 2  | AC119407 | AC119407 Pan trogl |
| c 19       | 78    | 7.8           | 175952 | 9  | AC069285 | AC069285 Homo sapi |
| c 20       | 66.4  | 6.6           | 57476  | 2  | AC107630 | AC107630 Homo sapi |
| c 21       | 52.2  | 5.2           | 125020 | 9  | AF429315 | AF429315 Homo sapi |
| c 22       | 47.4  | 4.7           | 7218   | 6  | I66494   | I66494 Sequence 14 |
| c 23       | 42.8  | 4.3           | 199654 | 2  | AC010898 | AC010898 Homo sapi |
| c 24       | 42    | 4.2           | 10732  | 6  | E32986   | E32986 Gene encodi |
| c 25       | 41.6  | 4.2           | 125020 | 9  | AF429315 | AF429315 Homo sapi |
| c 26       | 41.4  | 4.1           | 157197 | 2  | AC099042 | AC099042 Oryza sat |
| c 27       | 40.4  | 4.0           | 227856 | 2  | AC007908 | AC007908 Homo sapi |
| c 28       | 40.2  | 4.0           | 94682  | 2  | AC128917 | AC128917 Rattus no |
| c 29       | 40    | 4.0           | 165432 | 9  | AL512590 | AL512590 Human DNA |
| c 30       | 39.8  | 4.0           | 65581  | 9  | HSU91320 | HSU91320 Human DNA |
| c 31       | 39.8  | 4.0           | 210344 | 9  | AC007100 | AC007100 Homo sapi |
| c 32       | 39.4  | 3.9           | 69205  | 2  | AC120034 | AC120034 Homo sapi |
| c 33       | 39.4  | 3.9           | 160584 | 2  | AC022248 | AC022248 Homo sapi |
| c 34       | 39.2  | 3.9           | 230884 | 2  | AC079523 | AC079523 Mus muscu |
| c 35       | 38.8  | 3.9           | 78986  | 9  | AL354693 | AL354693 Human DNA |
| c 36       | 38.8  | 3.9           | 115666 | 2  | AC105744 | AC105744 Oryza sat |
| c 37       | 38.8  | 3.9           | 163282 | 2  | AC016438 | AC016438 Homo sapi |
| c 38       | 38.8  | 3.9           | 167553 | 2  | AC114938 | AC114938 Homo sapi |
| c 39       | 38.8  | 3.9           | 168435 | 2  | AL390764 | AL390764 Homo sapi |
| c 40       | 38.8  | 3.9           | 185304 | 2  | AC106269 | AC106269 Rattus no |
| c 41       | 38.8  | 3.9           | 191014 | 9  | AC113268 | AC113268 Papio cyn |
| c 42       | 38.6  | 3.9           | 146383 | 2  | AC116367 | AC116367 Oryza sat |
| c 43       | 38.6  | 3.9           | 170261 | 2  | AC008542 | AC008542 Homo sapi |
| c 44       | 38.6  | 3.9           | 173391 | 9  | AC090063 | AC090063 Homo sapi |
| c 45       | 38.6  | 3.9           | 189076 | 2  | AC023788 | AC023788 Homo sapi |

ALIGNMENTS

RESULT 1  
AC018558

LOCUS Homo sapiens chromosome 16 clone RP11-80F22, complete sequence.  
DEFINITION Homo sapiens chromosome 16 clone RP11-80F22, complete sequence.  
ACCESSION AC018558

VERSION AC018558.5 GI:16596530  
KEYWORDS HTG.  
SOURCE Homo sapiens.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

REFERENCE 1 (bases 1 to 151049)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 16  
JOURNAL Unpublished

|            |       |   |   |  |  |
|------------|-------|---|---|--|--|
| REFERENCE  |       | 2   | (bases 1 to 151049)   |  |  |
| AUTHORS    |       |   | DOE Joint Genome Institute.   |  |  |
| TITLE      |       |   | Direct Submission   |  |  |
| JOURNAL    |       |   | Submitted (14-DEC-1999) Production Sequencing Facility, DOE Joint<br>Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA   |  |  |
| REFERENCE  |       | 3   | (bases 1 to 151049)   |  |  |
| AUTHORS    |       |   | DOE Joint Genome Institute.   |  |  |
| TITLE      |       |   | Direct Submission   |  |  |
| JOURNAL    |       |   | Submitted (02-NOV-2001) Production Sequencing Facility, DOE Joint<br>Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA   |  |  |
| COMMENT    |       |   | On Nov 2, 2001 this sequence version replaced gi:9795566.<br><br>Sequence Quality Assessment:<br>This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.<br>All manually edited bases have been reduced to quality zero.<br>Quality levels above 40 are expected to have less than 1 error in 10,000 bp.<br>Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.<br>-----<br>Sequence Quality Assessment:<br>This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.<br>All manually edited bases have been reduced to quality zero.<br>Quality levels above 40 are expected to have less than 1 error in 10,000 bp.<br>Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.<br>----- |  |  |
| FEATURES   |       |   | Location/Qualifiers   |  |  |
| source     |       |   | 1..151049<br>/organism="Homo sapiens"<br>/db_xref="taxon:9606"<br>/chromosome="16"<br>/cloned="RP11-80F22"  |  |  |
| BASE COUNT |       |   | 46827 a 30783 c 30749 g 42690 t   |  |  |
| ORIGIN     |       |   |   |  |  |
|            |       |   | Query Match                100.0%; Score 1000; DB 9; Length 151049;<br>Best Local Similarity      100.0%; Pred. No. 1.3e-291;<br>Matches 1000; Conservative   0; Mismatches   0; Indels   0; Gaps   0   |  |  |
| Qy         | 1     | AGCCAGACTAGGAGTGGCAGAGAAGGGAAGGATGGTGAGGCACAGGTGCACCTCTA        | 60  |  |  |
| Db         | 70221 | AGCCAGACTAGGAGTGGCAGAGAAGGGAAGGATGGTGAGGCACAGGTGCACCTCTA        | 70280   |  |  |
| Qy         | 61    | CTGGTGCCCGCCAGACCAGACTGCATGCCAGGCTGCAGTCCCAAGGATACTCGGTGCGGG    | 120   |  |  |
| Db         | 70281 | CTGGTGCCCGCCAGACCAGACTGCATGCCAGGCTGCAGTCCCAAGGATACTCGGTGCGGG    | 70340   |  |  |
| Qy         | 121   | TCCTGTGCCCATAGCATCTTAGATCAGCTGCTGAGGCTGGAGGTTCTTCATTCTCTTG      | 180   |  |  |
| Db         | 70341 | TCCTGTGCCCATAGCATCTTAGATCAGCTGCTGAGGCTGGAGGTTCTTCATTCTCTTG      | 70400   |  |  |
| Qy         | 181   | AGCATCAGGGGTGTGTATCATTTTCCAAGGGTTTTTCAGACAATCCCTGTGTGACCCCTGSCA | 240   |  |  |
| Db         | 70401 | AGCATCAGGGGTGTGTATCATTTTCCAAGGGTTTTTCAGACAATCCCTGTGTGACCCCTGSCA | 70460   |  |  |
| Qy         | 241   | GGGGCGGTTTATCATGGCATCGGTCCATGGCCTTTGCCTCCAAGCAGCACCCAGCAATCC    | 300   |  |  |
| Db         | 70461 | GGGGCGGTTTATCATGGCATCGGTCCATGGCCTTTGCCTCCAAGCAGCACCCAGCAATCC    | 70520   |  |  |
| Qy         | 301   | CCATGCCCAAAATGACATTAATGTTTGTGTGGGCTCTTCTTGGGAAGCTCAGCTTCTC      | 360   |  |  |
| Db         | 70521 | CCATGCCCAAAATGACATTAATGTTTGTGTGGGCTCTTCTTGGGAAGCTCAGCTTCTC      | 70580   |  |  |
| Qy         | 361   | CTCTGTGTTCCCTTCCATCTTCCCCAAAACAGTACTTCTGGCATTCCTTGTTCACCAAC     | 420   |  |  |
| Db         | 70581 | CTCTGTGTTCCCTTCCATCTTCCCCAAAACAGTACTTCTGGCATTCCTTGTTCACCAAC     | 70640   |  |  |
| Qy         | 421   | AATGGGAAAACTGGGTCTGGAGACTTCAGAAACCACACTGTGAGGGCTTCGAGTCTTCCCT   | 480   |  |  |

Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

#### Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

#### FEATURES

Location/Qualifiers

1..191768  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-332P24"

BASE COUNT 57149 a 39592 c 40150 g 54877 t  
ORIGIN

Query Match 100.0%; Score 1000; DB 9; Length 191768;  
Best Local Similarity 100.0%; Pred. No. 1.3e-291;  
Matches 1000; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy 1 AGCCAGACTAGAGTGAGCCAGAGAGGGAAGGATGGTGAGGCACAGCGTGCACACTCTA 60  
|||||  
Db 19966 AGCCAGACTAGAGTGAGCCAGAGAGGGAAGGATGGTGAGGCACAGCGTGCACACTCTA 20025  
|||||

Qy 61 CTGGTGCCCCAGACCCAGACATGCATGCCAGGCTGCAGTCCAAAGGATACCTCGGTGCGGG 120  
|||||

Db 20026 CTGGTGCCCCAGACCCAGACATGCATGCCAGGCTGCAGTCCAAAGGATACCTCGGTGCGGG 20085  
|||||

Qy 121 TCCCTGTCCCCATAGCATCTTAGATCAGCTGCTGAGGCTGGAGTCTTCCATTCCTTG 180  
|||||

Db 20086 TCCCTGTCCCCATAGCATCTTAGATCAGCTGCTGAGGCTGGAGTCTTCCATTCCTTG 20145  
|||||

Qy 181 AGCATCAGGGTGTATCATCTTCCAAAGGTTTTCAGACAAATCCCTGTGAGCCCTGGCA 240  
|||||

Db 20146 AGCATCAGGGTGTATCATCTTCCAAAGGTTTTCAGACAAATCCCTGTGAGCCCTGGCA 20205  
|||||

Qy 241 GGGGCGGTTATCATGGCGATCGGTTCATGGCCCTTGCCTCCAAAGCAGACCCAGCAATCC 300  
|||||

Db 20206 GGGGCGGTTATCATGGCGATCGGTTCATGGCCCTTGCCTCCAAAGCAGACCCAGCAATCC 20265  
|||||

Qy 301 CCATGCCCAACCAATGCACATAATGTTTGTGGGGCCCTTTCTGGAAGCTCACCTTCTC 360  
|||||

Db 20266 CCATGCCCAACCAATGCACATAATGTTTGTGGGGCCCTTTCTGGAAGCTCACCTTCTC 20325  
|||||

Qy 361 CTCCTGTTTGGCCCTCCATCTTCCCAACACAGTACTTCTGCGCATCCCTTGTCCACCAC 420  
|||||

Db 20326 CTCCTGTTTGGCCCTCCATCTTCCCAACACAGTACTTCTGCGCATCCCTTGTCCACCAC 20385  
|||||

Qy 421 AATGGAAACTGGGTCTCGAGACTCAGAAACCACTGTGAGGCGCTCAGTCTTCCCT 480  
|||||

Db 20386 AATGGAAACTGGGTCTCGAGACTCAGAAACCACTGTGAGGCGCTCAGTCTTCCCT 20445  
|||||

Qy 481 GTCCTGGCTACAGGGCATGAATCAGAGAGAAAAGTCACTTCCACCTCTCGAAGGCTG 540  
|||||

Db 20446 GTCCTGGCTACAGGGCATGAATCAGAGAGAAAAGTCACTTCCACCTCTCGAAGGCTG 20505  
|||||

Qy 541 CCAGGCTCAGGCTGGGCACACTGAGGCTGACAGGGGCTTCTGAAGGCCAGAGAGATG 600  
|||||

Db 20506 CCAGGCTCAGGCTGGGCACACTGAGGCTGACAGGGGCTTCTGAAGGCCAGAGAGATG 20565  
|||||

Qy 601 GCCCGGACATAAGGCTGAAGCAACCTCTCTGAGCCAAAGATCTGTTTGTCTCTCTGA 660  
|||||

Db 20566 GCCCGGACATAAGGCTGAAGCAACCTCTCTGAGCCAAAGATCTGTTTGTCTCTCTGA 20625  
|||||

Qy 661 ATCTTAGTGGCTTCTTAAAGCGGGTGTGATCAGCCATGGGTATCAGAGACACTGGAGTC 720  
|||||

Db 20626 ATCTTAGTGGCTTCTTAAAGCGGGTGTGATCAGCCATGGGTATCAGAGACACTGGAGTC 20685  
Qy 721 CAGTAGCTCTAGGTGGGACACGGGCACAAATTTCACTTGCAGACCAGCTGCACGGAGTGG 780  
|||||

Db 20686 CAGTAGCTCTAGGTGGGACACGGGCACAAATTTCACTTGCAGACCAGCTGCACGGAGTGG 20745  
|||||

Qy 781 ATAAAGAGAGATCTCTGTGTGGGAATCTCCTTTGTGGTATCATCAGGGAGTGAAGTCTT 840  
|||||

Db 20746 ATAAAGAGAGATCTCTGTGTGGGAATCTCCTTTGTGGTATCATCAGGGAGTGAAGTCTT 20805  
|||||

Qy 841 TGTCTATAGCCATCATATCCAGCTTGTGTGATACCAATTCAGTGAAGCTGGACAACAGCTGG 900  
|||||

Db 20806 TGTCTATAGCCATCATATCCAGCTTGTGTGATACCAATTCAGTGAAGCTGGACAACAGCTGG 20865  
|||||

Qy 901 CACTGCTCAACAGCGCTTACCAAGACATCATGTTTTTTTTTTTTTTTCCACCAACCTG 960  
|||||

Db 20866 CACTGCTCAACAGCGCTTACCAAGACATCATGTTTTTTTTTTTTTTTCCACCAACCTG 20925  
|||||

Qy 961 GACCTGAATGGGATGTGGACACACATAGTCCAGAGGA 1000  
|||||

Db 20926 GACCTGAATGGGATGTGGACACACATAGTCCAGAGGA 20965  
|||||

#### RESULT 3

AC026186

LOCUS

DEFINITION Homo sapiens chromosome 3 clone RP11-303p8 map 3p, WORKING DRAFT  
SEQUENCE, 44 unordered pieces.

ACCESSION

AC026186

VERSION

AC026186.2 GI:8101183

KEYWORDS

HTG; HTGS\_PHASE1; HTGS\_DRAFT.

SOURCE

Homo sapiens.

ORGANISM

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.

REFERENCE

1 (Bases 1 to 167133)

Authors

Bao, J., Bao, Q., Bao, W., Bian, X., Cao, T., Chen, C., Chen, J., Ding, H.,

Dong, W., Fan, H., Feng, X., Guan, Q., Gu, X., Guo, D., He, L., Hu, S.,

Huang, F., Jin, Y., Kang, N., Li, C., Li, G., Li, J., Li, L.,

Li, S., Li, T., Liu, Y., Liu, B., Liu, Y., Li, W., Li, Y.,

Lu, J., Niu, Y., Qi, Q., Qi, X., Song, S., Sun, M., Sun, Y.,

Tao, R., Wang, H., Wang, J., Wang, L., Wang, L., Wang, R.,

Wang, X., Wang, X., Wang, Y., Wu, D., Wu, Q., Xie, F., Xuan, Z., Xue, Y.,

Yan, C., Yang, X., Yu, B., Zeng, Y., Zhang, G., Zhang, H., Zhang, H.,

Zhang, L., Zhang, M., Zhang, X., Zhang, X., Zhang, Y., Zhang, Y.,

Zhang, Z., Zhu, B., Yu, J. and Yang, H.

Chromosome 3p genomic sequence

Unpublished

2 (Bases 1 to 167133)

Authors

Chen, J., Hu, S., Dong, W., Wang, J., Zhang, Y., Zhang, H., Liu, B.,

Bao, W., Sun, Y., Wu, Q., Wang, H., Yang, X., Cheng, C., Wang, Y., Niu, Y.,

Qi, X., Li, T., Zhang, H., Liu, N., Wu, D., Yu, B., Fan, H., Liu, Y.,

Li, G., Li, C., Bao, Q., Bao, J., Wang, X., Song, L., Zhang, L., Guo, D.,

Huang, F., Zhang, G., Li, J., Bian, X., Zhang, M., Li, L., Feng, X., Yu, J.

and Yang, H.

Direct Submission

Submitted (21-MAR-2000) Human Genomic Center, Institute of

Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing

100101, P.R.China

On May 29, 2000 this sequence version replaced gi:7272003.

Center: Beijing Center

Website: http://hgsc.igtp.ac.cn

Contact: hgsc@igtp.ac.cn

----- Project Information

Center project name: 1% project

Center clone name: RP11-303p8

----- Summary Statistics

Sequencing vector: pUC18; 100% of reads

Chemistry: Dye-terminator; Et 55% of reads

Chemistry: Dye-terminator Big Dye; 45% of reads









searches of the EST database at TIGR  
 (http://www.tigr.org/tdb/hcd/hcd.html) and searches against a  
 peptide database. Repeats were identified using RepeatMasker (Smit,  
 A.F.A. and Green, P. unpublished,  
 http://ftp.genome.washington.edu/rm/RepeatMasker.html.

#### FEATURES

|        |                          |                          |
|--------|--------------------------|--------------------------|
| source | 1. .70048                |                          |
|        | /organism="Homo sapiens" |                          |
|        | /db_xref="taxon:9606"    |                          |
|        | /chromosome="16"         |                          |
|        | /map="16p13"             |                          |
|        | /clone="Cir987SK-3H8"    |                          |
|        | complement(574..637)     | /rpt_family="L1"         |
|        | /rpt_family="(GA)n"      | 18174..18287             |
|        | 869..1165                | /rpt_family="pTR5"       |
|        | /rpt_family="AluJb"      | 19188..19341             |
|        | complement(1240..1391)   | /rpt_family="LORI"       |
|        | /rpt_family="(TAAA)n"    | 19330..19455             |
|        | complement(1170..2642)   | /rpt_family="LORI"       |
|        | /rpt_family="MER42c"     | 19717..19757             |
|        | complement(2932..2952)   | /rpt_family="FLAM_C"     |
|        | /rpt_family="AT_rich"    | 19746..19829             |
|        | 2961..3072               | /rpt_family="LTRA"       |
|        | /rpt_family="AluJo"      | 19949..20122             |
|        | 3074..3104               | /rpt_family="L1PA13"     |
|        | /rpt_family="(CA)n"      | complement(21483..22408) |
|        | 3105..3282               | /rpt_family="L1MA8"      |
|        | 3782..4075               | complement(22419..22715) |
|        | /rpt_family="AluSc"      | /rpt_family="AluSc"      |
|        | 4492..4790               | complement(22722..22911) |
|        | /rpt_family="AluSg"      | /rpt_family="L1"         |
|        | complement(5321..5424)   | 22967..23190             |
|        | /rpt_family="L1MB3"      | /rpt_family="AluJb"      |
|        | 5412..5523               | complement(23191..23558) |
|        | /rpt_family="MER30"      | /rpt_family="L1"         |
|        | 5913..6204               | 23497..23739             |
|        | /rpt_family="AluSp"      | /rpt_family="MSTA"       |
|        | complement(6235..6448)   | complement(24058..24094) |
|        | /rpt_family="AluJo"      | /rpt_family="AT_rich"    |
|        | complement(6742..7035)   | complement(24204..24238) |
|        | /rpt_family="AluSg"      | /rpt_family="AT_rich"    |
|        | complement(7838..8142)   | complement(24753..24795) |
|        | /rpt_family="MER2"       | /rpt_family="AT_rich"    |
|        | 9144..9439               | complement(25016..25342) |
|        | /rpt_family="AluSp"      | /rpt_family="MER28"      |
|        | complement(10493..10786) | complement(25342..25388) |
|        | /rpt_family="AluSx"      | /rpt_family="TIGGER2"    |
|        | complement(10787..10914) | complement(25473..25554) |
|        | /rpt_family="L1MC3"      | /rpt_family="(TAAA)n"    |
|        | complement(11005..11226) | 25510..25633             |
|        | /rpt_family="AluYb8"     | /rpt_family="(CAT)n"     |
|        | 11968..12015             | complement(25745..25886) |
|        | /rpt_family="AT_rich"    | /rpt_family="(TAAA)n"    |
|        | 12300..12587             | complement(26103..26241) |
|        | /rpt_family="AluSg"      | /rpt_family="(TAAA)n"    |
|        | 13164..13453             | complement(26479..26591) |
|        | /rpt_family="AluSx"      | /rpt_family="(TGGAA)n"   |
|        | 13455..13499             | 26639..26765             |
|        | /rpt_family="(TA)n"      | /rpt_family="(CAT)n"     |
|        | complement(13574..13875) | complement(26664..26806) |
|        | /rpt_family="AluSx"      | /rpt_family="(TGAA)n"    |
|        | complement(14046..14346) | complement(27057..27192) |
|        | /rpt_family="AluSx"      | /rpt_family="(GAAAA)n"   |
|        | 14569..14867             | 27282..27400             |
|        | /rpt_family="AluSx"      | /rpt_family="(CAT)n"     |
|        | 15024..15327             | complement(27312..27450) |
|        | /rpt_family="AluSg"      | /rpt_family="(TAAA)n"    |
|        | complement(15560..15847) | complement(27787..27923) |
|        | /rpt_family="AluSc"      | /rpt_family="(TAAA)n"    |
|        | complement(16051..16357) | complement(28032..28156) |
|        | /rpt_family="AluSp"      | /rpt_family="(TGGAA)n"   |
|        | 16805..17007             | complement(28136..28272) |
|        |                          | /rpt_family="(TGAA)n"    |
|        |                          | complement(28465..28588) |
|        |                          | /rpt_family="(GAAAA)n"   |
|        |                          | complement(28899..29008) |
|        |                          | /rpt_family="(TGGAA)n"   |
|        |                          | 28939..29063             |
|        |                          | /rpt_family="(CAT)n"     |
|        |                          | complement(30056..30188) |
|        |                          | /rpt_family="FLAM_C"     |
|        |                          | complement(30634..30733) |
|        |                          | /rpt_family="MER30"      |

|               |                          |
|---------------|--------------------------|
| repeat_region | /rpt_family="AluSx"      |
| repeat_region | complement(17214..17823) |
| repeat_region | /rpt_family="L1"         |
| repeat_region | 18174..18287             |
| repeat_region | /rpt_family="pTR5"       |
| repeat_region | 19188..19341             |
| repeat_region | /rpt_family="LORI"       |
| repeat_region | 19330..19455             |
| repeat_region | /rpt_family="LORI"       |
| repeat_region | 19717..19757             |
| repeat_region | /rpt_family="FLAM_C"     |
| repeat_region | 19746..19829             |
| repeat_region | /rpt_family="LTRA"       |
| repeat_region | 19949..20122             |
| repeat_region | /rpt_family="L1PA13"     |
| repeat_region | complement(21483..22408) |
| repeat_region | /rpt_family="L1MA8"      |
| repeat_region | complement(22419..22715) |
| repeat_region | /rpt_family="AluSc"      |
| repeat_region | complement(22722..22911) |
| repeat_region | /rpt_family="L1"         |
| repeat_region | 22967..23190             |
| repeat_region | /rpt_family="AluJb"      |
| repeat_region | complement(23191..23558) |
| repeat_region | /rpt_family="L1"         |
| repeat_region | 23497..23739             |
| repeat_region | /rpt_family="MSTA"       |
| repeat_region | complement(24058..24094) |
| repeat_region | /rpt_family="AT_rich"    |
| repeat_region | complement(24204..24238) |
| repeat_region | /rpt_family="AT_rich"    |
| repeat_region | complement(24753..24795) |
| repeat_region | /rpt_family="AT_rich"    |
| repeat_region | complement(25016..25342) |
| repeat_region | /rpt_family="MER28"      |
| repeat_region | complement(25342..25388) |
| repeat_region | /rpt_family="TIGGER2"    |
| repeat_region | complement(25473..25554) |
| repeat_region | /rpt_family="(TAAA)n"    |
| repeat_region | 25510..25633             |
| repeat_region | /rpt_family="(CAT)n"     |
| repeat_region | complement(25745..25886) |
| repeat_region | /rpt_family="(TAAA)n"    |
| repeat_region | complement(26103..26241) |
| repeat_region | /rpt_family="(TAAA)n"    |
| repeat_region | complement(26479..26591) |
| repeat_region | /rpt_family="(TGGAA)n"   |
| repeat_region | 26639..26765             |
| repeat_region | /rpt_family="(CAT)n"     |
| repeat_region | complement(26664..26806) |
| repeat_region | /rpt_family="(TGAA)n"    |
| repeat_region | complement(27057..27192) |
| repeat_region | /rpt_family="(GAAAA)n"   |
| repeat_region | 27282..27400             |
| repeat_region | /rpt_family="(CAT)n"     |
| repeat_region | complement(27312..27450) |
| repeat_region | /rpt_family="(TAAA)n"    |
| repeat_region | complement(27787..27923) |
| repeat_region | /rpt_family="(TAAA)n"    |
| repeat_region | complement(28032..28156) |
| repeat_region | /rpt_family="(TGGAA)n"   |
| repeat_region | complement(28136..28272) |
| repeat_region | /rpt_family="(TGAA)n"    |
| repeat_region | complement(28465..28588) |
| repeat_region | /rpt_family="(GAAAA)n"   |
| repeat_region | complement(28899..29008) |
| repeat_region | /rpt_family="(TGGAA)n"   |
| repeat_region | 28939..29063             |
| repeat_region | /rpt_family="(CAT)n"     |
| repeat_region | complement(30056..30188) |
| repeat_region | /rpt_family="FLAM_C"     |
| repeat_region | complement(30634..30733) |
| repeat_region | /rpt_family="MER30"      |

repeat\_region complement(30864..31123)  
repeat\_region /rpt\_family="AluJo"  
repeat\_region 32213..32238  
repeat\_region /rpt\_family="GC\_rich"  
repeat\_region 32353..32518  
repeat\_region /rpt\_family="(CGG)n"  
repeat\_region complement(32639..32781)  
repeat\_region /rpt\_family="(TA)n"  
repeat\_region complement(32728..32877)  
repeat\_region /rpt\_family="(TAA)n"  
repeat\_region 33451..33494  
repeat\_region /rpt\_family="(CA)n"  
repeat\_region 34354..34767  
repeat\_region /rpt\_family="L1"  
repeat\_region 34786..34917  
repeat\_region /rpt\_family="FLAM\_A"  
repeat\_region 35026..35342  
repeat\_region /rpt\_family="L1"  
repeat\_region complement(35555..36427)  
repeat\_region /rpt\_family="L1PAL6"  
repeat\_region complement(36279..36847)

Query Match 26.9%; Score 269.4; DB 9; Length 70048;  
Best Local Similarity 76.5%; Pred. No. 4.3e-70;  
Matches 368; Conservative 0; Mismatches 86; Indels 27; Gaps 2;  
QY 525 CACCTCTGAAGCGTCCAGCGCTGCGAGCGTGGCACACTGAGCGTGCACAGGGCCCTTCG 584  
Db 51193 CCCACGGAGAGACCGCTAGGCTGAGGCTTGTACGTGAGCGCTGGCAGGGCCCTTCGG 51252  
QY 585 AAGGCCAGGAGGATGGCCGGGACATAGGCTGAAGCAACCTGCTGTAGCCAAAGATCT 644  
Db 51253 AAGCAAGGGAAGATGGCCAGGACACAAAGCTGAAGCAACCCATCTGAGCCAAAGATCT 51312  
QY 645 GTTGTGTCTCTCTGATCTTGTAGTGGCTTCTAAAGGCGGGTGTGATCAGCCATGGGTAT 704  
Db 51313 GTTGTGTCTCTCTGATCTTGTAGTGGCTTCTAAAGGCGGGTGTGATCAGCCATGGGTAT 51372  
QY 705 CAGACACTGGAGTCCAGTACTGCTAGTGGG-----ACACGGGC 746  
Db 51373 CAGACACTGGAGTCCAGTACTGCGCTGGCGGTGGCCAGGAGCGCTGACTACTGCACAGT 51432  
QY 747 ACAATTTACTTGCAGACAGCTGCACGGAGTGGATAAGAGAGAGTCTGTGTGGGAAT 806  
Db 51433 GTGATTTACTGGCAGGCCAAGCTGCAGGGAGTGGATAAGAGAGAGTCTGTGTGGGAAT 51492  
QY 807 CTCCTTGTGTGATCATCAGGAGGTGAAGTCTTTGTCTATACCTCATATCCAGCTTGTG 866  
Db 51493 CTCCTTGTGTGATCATCAGGAGGTGAAGTCTTTGTCTATACCTCATATCCAGCTTGTG 51552  
QY 867 TGATACCAATTCAGTGAAGCTGGCAAGCTGGCAGCTGCTCAACAGCGCCTACCAAGAC 926  
Db 51553 GATAGCAATTCAGTGAAGCTGGCAAGCTGGCAGCTGCTCAACAGCGCCTACCAAGAT 51612  
QY 927 ATCATGTTTTTTTTTTTTTTTCCACCAACCTGACCTGTAATGGGATGTGGACACACA 986  
Db 51613 ACCAGG-----TTTCTTTCCACACAGCTGGCGCCTGGATGGGATGTGAACATCTA 51663  
QY 987 T 987  
Db 51664 T 51664

RESULT 7  
AC020761/c 129290 bp DNA linear HTG 25-APR-2001  
LOCUS Homo sapiens chromosome 16 clone RP11-151M15, WORKING DRAFT  
DEFINITION SEQUENCE, 20 unordered pieces.  
AC020761  
VERSION AC020761.5 GI:9965531  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT; HTGS\_ACTIVEFIN.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
COMMENT

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.  
1 (bases 1 to 129290)  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 16  
Unpublished  
2 (bases 1 to 129290)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (09-JAN-2000) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Sep 3, 2000 this sequence version replaced gi:9098615.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov

Project Information  
Center Project Name: 0  
Center clone name: RPCI-11\_151M15  
-----

Summary Statistics  
Consensus quality: 96142 bases at least Q40  
Consensus quality: 106591 bases at least Q30  
Consensus quality: 112022 bases at least Q20  
Estimated insert size: 161300; agarose-fp estimation  
Estimated insert size: 127390; sum-of-contigs estimation  
Quality coverage: 4.28 in Q20 bases; agarose-fp estimation  
Quality coverage: 5.42 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 20 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 1043: contig of 1043 bp in length  
1044 1143: gap of unknown length  
1144 2482: contig of 1339 bp in length  
2483 2582: gap of unknown length  
2583 3593: contig of 1011 bp in length  
3594 3693: gap of unknown length  
3694 4803: contig of 1110 bp in length  
4804 4903: gap of unknown length  
4904 6120: contig of 1217 bp in length  
6121 6220: gap of unknown length  
6221 7568: contig of 1348 bp in length  
7569 9284: contig of 1616 bp in length  
9285 9384: gap of unknown length  
9385 10398: contig of 1014 bp in length  
10399 10498: gap of unknown length  
10499 12201: contig of 1703 bp in length  
12202 12301: gap of unknown length  
12302 13971: contig of 1670 bp in length  
13972 14071: gap of unknown length  
14072 15687: contig of 1616 bp in length  
15688 15787: gap of unknown length  
15788 16900: contig of 1113 bp in length  
16901 17000: gap of unknown length  
17001 19643: contig of 2643 bp in length  
19644 19743: gap of unknown length  
19744 22031: contig of 2288 bp in length  
22032 22131: gap of unknown length  
22132 22999: contig of 5168 bp in length  
22999 27300: gap of unknown length  
27300 33138: contig of 5739 bp in length  
33139 33238: gap of unknown length  
33239 41651: contig of 8413 bp in length  
41652 41751: gap of unknown length  
41752 47053: contig of 5302 bp in length  
47054 47153: gap of unknown length

```
* 47154 88215: contig of 41062 bp in length
* 88316 88315: gap of unknown length
* 88316 129290: contig of 40975 bp in length.
FEATURES
  Location/Qualifiers
    1..129290
      /organism="Homo sapiens"
      /db_xref="taxon:9606"
      /chromosome="16"
      /clone="RP11-151M15"
      /clone_lib="RPCI human BAC library 11"
BASE COUNT 34282 a 28372 c 28553 g 36024 t 2059 others
ORIGIN

Query Match 26.9%; Score 269.4; DB 2; Length 129290;
Best Local Similarity 76.5%; Pred. No. 4.3e-70;
Matches 368; Conservative 0; Mismatches 86; Indels 27; Gaps 2;

Qy 525 CACCTCTGAAGGCTGCAGCGCTGAGGCTTGGCACACTGAGGTGACAGGGGCTTCTGTG 584
Db 69206 CCCACGGAGAGACCGCTAGGCTTGGTGAAGCTGAGGCTGGCAGGGAGCTTCGG 69147

Qy 585 AAGCCAGAGAGATGGCCGGGACATAAGGCTGAAGCACTGTCTGAGCCAAAGATCT 644
Db 69146 AAGCAAGGAGATGGCCAGGACACAAAGCTGAAGCAACCCTATGAGCCAAAGATCT 69087

Qy 645 GTTGTGTCTCTCCTGAATCTTAGTGGCTTCTAAAGGGGGGTGTGATCAGCCATGGGTAT 704
Db 69086 GTTTAGGTCTCTCTGAATCTCAGCAGCCTTCCCAAGGAAGCATGATTACCCATGGGTAT 69027

Qy 705 CAGACACTGAGTCCAGTACGTAGTGGG-----ACACGGGC 746
Db 69026 CAGACACTGAGGCTGGCAGCTGGCGGTGGCCAGAGGCTGACTACTGCACAGCT 68967

Qy 747 ACAATTTCACTTGACAGCACCTGCACGAGTGGATGAAGAGAGATCTCTGTGGGAAT 806
Db 68966 GTGATTTCACTTGGCAGGCCCACTCAGGAGTGGATGAAGAGAGCTCTGTGGGAAT 68907

Qy 807 CTCTCTGGTGGATCAGGAGGTGAAGTCTTTGTTCATAGCTCATATCAGCTTGTG 866
Db 68906 CTCTCTGGTGGATCATCGAAGAGTGAAGTCTTCTTCATAGCTCAACCCCAATTTGTG 68847

Qy 867 TGATACCAATTCAGTGAAGCTGAACAGCTGGCAGTGTCTCAACAGGCTTACCAAGAC 926
Db 68846 GGATAGCAATTCAGTGAAGCTGGGCAAGCTGGCAGTGTCTCAACAGGCTTCCCAAGAT 68787

Qy 927 ATCATGTTTTTTTTTTTTTTTCCACCAACCTGACCTGATGGGATGTGGACACACA 986
Db 68786 ACCAGG-----TTTCTTTCCACCACACAGCTGGGCTGGATTGGGATGTGAACATCTA 68736

Qy 987 T 987
Db 68735 T 68735

RESULT 8
AC114938/c
LOCUS
DEFINITION Homo sapiens chromosome 5 clone CTD-2244D9, WORKING DRAFT SEQUENCE,
3 unordered pieces.
AC114938
VERSION AC114938.1 GI:19424399
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
  Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
  1 (bases 1 to 167553)
  DOE Joint Genome Institute.
  TITLE Sequencing of Human Chromosome 5
  JOURNAL Unpublished
  REFERENCE 2 (bases 1 to 167553)
  DOE Joint Genome Institute.
  AUTHORS Direct Submission
  TITLE
```

Submitted (14-MAR-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Project Information  
Center Project Name: 721326  
Center clone name: C1TB-HI\_2244D9  
-----  
Summary Statistics  
Consensus quality: 163638 bases at least Q40  
Consensus quality: 165018 bases at least Q30  
Consensus quality: 166006 bases at least Q20  
Estimated insert size: 150000; agarose-fp estimation  
Estimated insert size: 167353; sum-of-contigs estimation  
Quality coverage: 10.56 in Q20 bases; agarose-fp estimation  
Quality coverage: 9.46 in Q20 bases; sum-of-contigs estimation.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 3 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 31475: contig of 31475 bp in length  
\* 31476 31575: gap of unknown length  
\* 31576 82160: contig of 50585 bp in length  
\* 82161 82260: gap of unknown length  
\* 82261 167553: contig of 85293 bp in length.  
FEATURES  
 Location/Qualifiers  
 1..167553  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="5"  
 /clone="CTD-2244D9"  
 /clone\_lib="Caltech human BAC library D"  
BASE COUNT 48962 a 35722 c 32192 g 50416 t 261 others  
ORIGIN

Query Match 26.9%; Score 269.4; DB 2; Length 167553;  
Best Local Similarity 76.5%; Pred. No. 4.3e-70;  
Matches 368; Conservative 0; Mismatches 86; Indels 27; Gaps 2;

Qy 525 CACCTCTGAAGGCTGCAGCGCTTGGCACACTGAGGTGACAGGGGCTTCTGTG 584  
Db 63442 CCCACGGAGAGACCGCTAGGCTTGGTACAGTGGGCTGGCAGGGACCTTCGG 63383

Qy 585 AAGCCAGAGAGATGGCCGGGACATAAGGCTGAAGCACTGTCTGAGCCAAAGATCT 644  
Db 63382 AAGCAAGGAGAGATGGCCAAAGGACCAAGGCTGAAGCAACCCTATCTGAGCCAAAGATCT 63323

Qy 645 GTTGTGTCTCTCCTGAATCTTAGTGGCTTCTAAAGGGGGGTGTGATCAGCCATGGGTAT 704  
Db 63322 GTTTAGGTCTCTCTGAATCTCAGCAGCCTTCCCAAGGAAGCATGATTACCCATGGGTAT 63263

Qy 705 CAGACACTGAGTCCAGTACGTAGTGGG-----ACACGGGC 746  
Db 63262 CAGACACTGGAGCTGGCAGCTGGCGGTGGCCAGGAGGCTGACTACTGCACAGCT 63203

Qy 747 ACAATTTCACTTGACAGCACCTGCACGAGTGGATGAAGAGAGATCTCTGTGGGAAT 806  
Db 63202 GTGATTTCACTTGGCAGGCCCAACTGACGGGAGTGGATGAAGAGAGAGCTCTGTGGGAAT 63143

Qy 807 CTCCTTTGGTGGATCATCAGGAGGTGAAGTCTTTGTATAGCTCATATCAGCTTGTG 866  
Db 63142 CTCTCTCGTGGATCATCGAAGAGTGAAGTCTTCTTCATAGCTCAACCCCAATTTGTG 63083

Qy 867 TGATACCAATTCAGTGAAGCTGGAACAAAGCTGGCACTGTCTCAACAGGCTTACCAAGAC 926  
Db 63082 GGATAGCAATTCAGTGAAGCTGGGACAAAGCTGGCACTGTCTCAACAGGCTTCCCAAGAT 63023





```
repeat_region 4920..4970 /rpt_family="MER2_type"
repeat_region 4974..5104 /rpt_family="MER2_type"
repeat_region 5116..5219 /rpt_family="L1"
repeat_region 5354..5416 /rpt_family="MER2_type"
misc_feature 5602..5955 /note="similar to EST AI806989 (NID:g5393555) wf24h01.x1"
repeat_region 5824..5887 /rpt_family="MER103"
misc_feature 6368..6806 /note="similar to EST AI187426 (NID:g3738064) qf31d08.x1"
repeat_region 6368..6520 /rpt_family="Alu"
repeat_region 6886..7074 /rpt_family="MER1_type"
repeat_region 7184..7418

Query Match 26.9%; Score 269.4; DB 9; Length 214025;
Best Local Similarity 76.5%; Pred. No. 4.3e-70;
Matches 368; Conservative 0; Mismatches 86; Indels 27; Gaps 2;

QY 525 CACCTCTGAAGCTGCCAGCTCAGGGCTTGGACACTGAGGCTGACAGGGGCGCTTCGTG 584
Db 27450 CCCCACGGAGAGACCGCTAGGCTCAGGGCTTGGTACAGTGAAGCTGGCAGGGCGCTTCGG 27509
QY 585 AAGGCAGAGGAGATGGCCCGGACATAGGCTGAAGCAACCTGCTGAGCCAAAGATCT 644
Db 27510 AAGGCAGGGAAGATGGCCCAAGGACACAAAGCTGAAGCAACCCATCTGAGCCAAAGATCT 27569
QY 645 GTTGTGTCTCCTCTGAATCTTAGTGGCCCTTCTAAAGCGGGGTGTGATCAGCCATGGGTAT 704
Db 27570 GTTGTGTCTCCTCTGAATCTCAGCAGCCCTTGCAGGAAGGATGATTTACCCATGGGTAT 27629
QY 705 CAGAGCACTGAGTCCAGTCACTGCTAGGTGGG-----ACAGGGGC 746
Db 27630 CAGAGCACTGAGGCTGCAGCTGGCGGTGGCGGAGGAGGCGCTGACTACTGCACAGCT 27689
QY 747 ACAATTTCACTTGCACAGCAGCTGCAGCGAGTGGATGAAGAGAGAGTCTGTGTGGGAAT 806
Db 27690 GTGATTTCACTGCGAGGCCAATCTCAGGGAGTGGATGAAGAGAGAGCTCTGTGTGGGAAT 27749
QY 807 CTCCTTTGGTGGATCAGGAGGTGAAGTCTTTGTATAGCCCTCATATCCAGCTTGTG 866
Db 27750 CTCCTCGTGGATCATCAGAGAGGTGAAGTCTTCTCATAGCCCTCAACCCCAATTTGTG 27809
QY 867 TGATACCAATTCAGTGAAGCTGGACAAAGCTGGCACTGCTCAACAGGCGCTTACCAAGAC 926
Db 27810 GGATAGCAATTCAGTGAAGCTGGACAAAGCTGGCACTGCTCAACAGGCGCTTCCCAAGAT 27869
QY 927 ATCATGTTTTTTTTTTTTTTTCCACCAACCTGGACCTGAATGGGATGGGACACACA 986
Db 27870 ACCAGG-----TTTCTTCCACCAAGCTGGGCGCTGGATGGGATGTGAACATCTA 27920
QY 987 T 987
Db 27921 T 27921

RESULT 11
AC007908/c
LOCUS AC007908 227856 bp DNA linear HTG 25-APR-2001
DEFINITION Homo sapiens chromosome 16 clone RP11-499D5, WORKING DRAFT
SEQUENCE, 30 unordered pieces.
ACCESSION AC007908
VERSION AC007908.3 GI:9094205
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
```

```
REFERENCE 1 (bases 1 to 227856)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 227856)
AUTHORS Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
Bryant,J., Tesmer,J., Meinkne,L., Longmire,J., White,S., Tatum,O.,
Campbell,C., Fawcett,J., Malthbie,M., Bussod,M., Sutherland,R.,
McMurry,K., Han,C. and Deaven,L.
JOURNAL Direct Submission
TITLE Submitted (24-JUN-1999) Center for Human Genome Studies, DOE Joint
JOURNAL Genome Institute, Los Alamos National Laboratory, MS M888, Los
Alamos, NM 87545, USA
COMMENT On Jul 13, 2000 this sequence version replaced gi:7211873.
-----Genome Center
Center: Joint Genome Institute
Center Code: JGI
Web site: http://www.jgi.doe.gov
-----
Project Information
Center Project Name: 0
Center clone name: RPCI-11_499D5
-----
Summary Statistics
Consensus quality: 184734 bases at least Q40
Consensus quality: 201642 bases at least Q30
Consensus quality: 209314 bases at least Q20
Estimated insert size: 218590; agarose-fp estimation
Quality coverage: 4.43 in Q20 bases; agarose-fp estimation
Quality coverage: 4.3 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
* 1259: contig of 1259 bp in length
* 1359: gap of unknown length
* 1360: contig of 1096 bp in length
* 2455: gap of unknown length
* 2555: gap of unknown length
* 3691: contig of 1136 bp in length
* 3692: gap of unknown length
* 3791: gap of unknown length
* 3792: contig of 1014 bp in length
* 4805: gap of unknown length
* 4905: contig of 1033 bp in length
* 5938: gap of unknown length
* 6038: gap of unknown length
* 7051: contig of 1013 bp in length
* 7151: gap of unknown length
* 8246: contig of 1095 bp in length
* 8346: gap of unknown length
* 8347: contig of 1647 bp in length
* 9993: gap of unknown length
* 10093: contig of 1282 bp in length
* 11375: gap of unknown length
* 11475: gap of unknown length
* 12513: contig of 1038 bp in length
* 12514: gap of unknown length
* 12614: contig of 1032 bp in length
* 13645: gap of unknown length
* 13646: contig of 1507 bp in length
* 13746: gap of unknown length
* 15252: contig of 1221 bp in length
* 15253: gap of unknown length
* 15353: contig of 1532 bp in length
* 16574: gap of unknown length
* 16674: contig of 1608 bp in length
* 18206: gap of unknown length
* 18306: contig of 1608 bp in length
* 19914: gap of unknown length
* 20014: contig of 2466 bp in length
* 22480: gap of unknown length
* 22580: contig of 2714 bp in length
```





\* 15788 16900: contig of 1113 bp in length  
\* 16901 17000: gap of unknown length  
\* 17001 19643: contig of 2643 bp in length  
\* 19644 19743: gap of unknown length  
\* 19744 22031: contig of 2288 bp in length  
\* 22032 22131: gap of unknown length  
\* 22132 27299: contig of 5168 bp in length  
\* 27300 33138: contig of 5739 bp in length  
\* 33139 33238: gap of unknown length  
\* 33239 41651: contig of 8413 bp in length  
\* 41652 47053: contig of 5302 bp in length  
\* 47054 47153: gap of unknown length  
\* 47154 88215: contig of 41062 bp in length  
\* 88216 88315: gap of unknown length  
\* 88316 129290: contig of 40975 bp in length.  
FEATURES  
    Location/Qualifiers  
        1..129290  
        /organism="Homo sapiens"  
        /db\_xref="taxon:9606"  
        /chromosome="16"  
        /clone="RP11-151M15"  
        /clone\_lib="RP11-151M15"  
BASE COUNT 34282 a 28372 c 28553 g 36024 t 2059 others  
ORIGIN  
Query Match 23.2%; Score 232.2; DB 2; Length 129290;  
Best Local Similarity 75.1%; Pred No. 8.1e-59;  
Matches 325; Conservative 0; Mismatches 86; Indels 20; Gaps 2;  
QY 525 CACCTCTGAAGGCTGCCAGCGTTCAGGCTTGCGACACTGAGGCTGACAGGGCCCTTCG 584  
Db 41178 CCCACGGAGAGACCACTAGGCTCAGGCTTGCTACAGTGAGGCTGGCAGGCGCCCTCGG 41237  
QY 585 AGGCCAGAGGATGGCCCGGACATAGGCTGAAGCAACCTGCTGAGCCAAAGATCT 644  
Db 41238 AAGGCAAGGGAAGTGGCCCAAGGACACAAAGGCTGAAGCAACCCATCTGAGCCAAAGATCT 41297  
QY 645 GTTGTGTCCTCTGAATCTTAGTGCCCTCTAAAGCGCGGTGTGATGACGCGATGGTAT 704  
Db 41298 GTTGTGTCCTCTGAATCTTAGTGCCCTCTAAAGCGCGGTGTGATGACGCGATGGTAT 41357  
QY 705 CAGAGACACTGGAGTCCAGTAGTGTCTAGGTGGG-----ACACGGG 745  
Db 41358 CAGAGACACTGGAGTCCAGTAGTGTCTAGGTGGG-----ACACGGG 745  
QY 746 CACATTTCACTTGCAGACAGCTGCACCGAGTGGATTAAGAGAGAGTCTGTGTGGGAA 805  
Db 41418 TGTGATTTCACTTGCAGACAGCTGCACCGAGTGGATTAAGAGAGAGTCTGTGTGGGAA 41477  
QY 806 TCTCCTTTGTGATCAGGAGGTGAAGTCTTTGTTCATAGCCCTCATATCCAGCTTCT 865  
Db 41478 TCTCCTCGTGTGATCATGAAGAGTGAAGTCTTCTTATAGCCCTTACCCCAATTTGT 41537  
QY 866 GTGATACCAATTCAGTGAAGCTGGAACAGCTGGCAGCTGCTCAACAGGCGCTTACCAAGA 925  
Db 41538 G-GATGACCAATTCAGTGAAGCTGGAACAGCTGGCAGCTGCTTGTTCACGAGCGCTTCCAAGA 41596  
QY 926 CATCATGTTTTTT 938  
Db 41597 TACCAGTTTCTTT 41609  
RESULT 13  
AC107630 AC107630 57476 bp DNA linear HTG 24-JAN-2002  
LOCUS Homo sapiens chromosome 17 clone CTD-2320G7 map 17, LOW-PASS  
DEFINITION SEQUENCE SAMPLING.  
ACCESSION AC107630  
VERSION AC107630.1 GI:18308327  
KEYWORDS HTG; HTGS\_PHASE0.  
SOURCE Homo sapiens.

ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE  
    1 (bases 1 to 57476)  
AUTHORS Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
TITLE Homo sapiens chromosome 17, clone CTD-2320G7  
JOURNAL Unpublished  
AUTHORS  
    2 (bases 1 to 57476)  
REFERENCE  
    1 (bases 1 to 57476)  
AUTHORS Birren, B., Linton, L., Nusbaum, C., Lander, E., Allen, N.,  
Anderson, S., Barna, N., Bastien, V., Boguslavsky, L., Boukhgalter, B.,  
Brown, A., Camarata, J., Campopiano, A., Chang, J., Chazaro, B.,  
Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cook, A.,  
Cooke, P., Dearellano, K., Dewar, K., Diaz, J. S., Dodge, S., Faro, S.,  
Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, N.,  
Ginde, S., Gord, S., Goyette, M., Graham, L., Grand-Pierre, N.,  
Hagos, B., Horton, L., Hulme, W., Iliev, I., Johnson, R., Jones, C.,  
Kamat, A., Karatas, A., Kellis, C., LaRoque, K., Lamazares, R.,  
Landers, T., Lehoczy, J., Levine, R., Liu, G., Maclean, C.,  
Macdonald, P., Major, J., Marquis, N., Matthews, C., McCarthy, M.,  
McEwan, P., McKernan, K., Meldrum, J., Meneus, L., Mihova, T.,  
Mlenga, V., Murphy, T., Naylor, J., Nguyen, C., Nicol, R., Norbu, C.,  
Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, J.,  
Peterson, K., Phunkhang, P., Pierre, N., Pollara, V., Raymond, C.,  
Retta, R., Rieback, M., Riley, R., Rise, C., Rogov, P., Roman, J.,  
Rosetti, M., Roy, A., Santos, R., Schauer, S., Schupback, R., Seaman, S.,  
Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N.,  
Strauss, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J.,  
Topham, K., Travers, M., Travis, N., Trigilio, J., Vassiliev, H.,  
Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G.,  
Zainoun, J., Zembek, L., Zimmer, A. and Zody, M.  
Direct Submission  
Submitted (24-JAN-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: tl2868  
Center clone name: 2320\_G\_7  
-----  
\* NOTE: This record contains 70 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows  
\* overlap relationships among clones to be deduced.  
\* However, it should not be assumed that this clone  
\* will be sequenced to completion. In the event that  
\* the record is updated, the accession number will  
\* be preserved.  
\* 1 699: contig of 699 bp in length  
\* 700 799: gap of 100 bp  
\* 800 1503: contig of 704 bp in length  
\* 1504 1603: gap of 100 bp  
\* 1604 2312: contig of 709 bp in length  
\* 2313 2412: gap of 100 bp  
\* 2413 3133: contig of 721 bp in length  
\* 3134 3233: gap of 100 bp  
\* 3234 3965: contig of 732 bp in length  
\* 3966 4065: gap of 100 bp  
\* 4066 4773: contig of 708 bp in length  
\* 4774 4873: gap of 100 bp  
\* 4874 5549: contig of 676 bp in length  
\* 5550 5649: gap of 100 bp  
\* 5650 6370: contig of 721 bp in length  
\* 6371 6470: gap of 100 bp  
\* 6471 7189: contig of 719 bp in length  
\*  
TITLE  
JOURNAL  
COMMENT



```

RESULT 14
AC092447
LOCUS      176547 bp      DNA      linear      PRI 27-FEB-2002
DEFINITION Homo sapiens BAC clone RP11-760D2 from 7, complete sequence.
ACCESSION  AC092447
VERSION     AC092447.5  GI:18873962
KEYWORDS   HTG.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 176547)
AUTHORS   Sulston, J.E. and Waterston, R.
TITLE     Toward a complete human genome sequence
JOURNAL   Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE   99063792
PUBMED    9847074
REFERENCE  2 (bases 1 to 176547)
AUTHORS   Armstrong, J. and Cotton, M.
TITLE     The sequence of Homo sapiens BAC clone RP11-760D2
JOURNAL   Unpublished (2001)
REFERENCE  3 (bases 1 to 176547)
AUTHORS   Waterston, R.H.
TITLE     Direct Submission
JOURNAL   Submitted (04-JUL-2001) Genome Sequencing Center, Washington
          University School of Medicine, 4444 Forest Park Parkway, St. Louis,
          MO 63108, USA
REFERENCE  4 (bases 1 to 176547)
AUTHORS   Waterston, R.H.
TITLE     Direct Submission
JOURNAL   Submitted (27-FEB-2002) Department of Genetics, Washington
          University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
          On Feb 23, 2002 this sequence version replaced gi:16259201.
COMMENT   ----- Genome Center
          Center: Washington University Genome Sequencing Center
          Center code: WUGSC
          Web site: http://genome.wustl.edu/gsc
          Contact: sapiens@watson.wustl.edu
          ----- Summary Statistics
          Center project name: H_NH0760D02
          -----

```

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

#### MAPPING INFORMATION:

The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/CTB/CHR7>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

#### SOURCE INFORMATION:

The RPCI-11 human BAC library was made from the blood of one male

donor, as described by Osoegawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tatenno, M., Catanese, J. J. and de Jong, P. J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org> VECTOR: pBACe3.6

#### NEIGHBORING SEQUENCE INFORMATION:

The clone sequenced to the left is RP13-492C18; the clone sequenced to the right is RP11-10F11, 2000 bp overlap. Actual start of this clone is at base position 1 of RP11-760D2; actual end is at base position 17615 of RP11-10F11.

The region from 20341 to 20385 is covered only by a PCR product from clone DNA. An unresolved simple sequence repeats exists between 84203 and 84724. Polymorphisms exist between AC024199 and AC092447. Data from AC024199 and AC092423 was used to finish AC092447.

#### FEATURES

| source       | Location/Qualifiers                                     |
|--------------|---|
| 1. 176547    | /organism="Homo sapiens"                                |
|              | /db_xref="taxon:9606"                                   |
|              | /map="7"  |
|              | /chromosome="7"   |
|              | /clone_lib="RPCI-11"                                    |
| 5. 1938      | /clone="RP11-760D2"                                     |
|              | /rpt_family="L1"  |
| 2210..2483   | /rpt_family="Alu"                                       |
| 2799..3012   | /note="match to EST BF894500 (NID:g12285959)"           |
|              | /note="match to EST BF895059 (NID:g12286518)"           |
| 2799..2962   | /rpt_family="L1"  |
| 3092..3388   | /rpt_family="L2"  |
| 3419..3536   | /rpt_family="L1"  |
| 3537..3794   | /rpt_family="Alu"                                       |
| 3795..4263   | /rpt_family="L1"  |
| 4259..4542   | /rpt_family="L1"  |
| 4548..4748   | /rpt_family="L1"  |
| 4749..5046   | /rpt_family="Alu"                                       |
| 5047..5267   | /rpt_family="L1"  |
| 5268..5560   | /rpt_family="Alu"                                       |
| 5561..6091   | /rpt_family="L1"  |
| 6528..6825   | /rpt_family="Alu"                                       |
| 8380..8686   | /rpt_family="Alu"                                       |
| 8924..9225   | /rpt_family="Alu"                                       |
| 9838..10125  | /rpt_family="Alu"                                       |
| 10336..10600 | /rpt_family="Alu"                                       |
| 11975..12282 | /rpt_family="Alu"                                       |
| 12457..12765 | /rpt_family="Alu"                                       |
| 14411..14448 | /note="match to EST AI334287 (NID:g4070846) qq25e08.xl" |
|              | /note="similar to Homo sapiens EST BF979990"            |

```
misc_feature (NID:g12347205)"
14412. .14455
/Note="similar to Homo sapiens EST BE879245
(NID:g10328021)"
14412. .14448
/Note="similar to Homo sapiens EST AL520426
(NID:g12783919)"
14530. .14608
/rpt_family="L1"
14609. .14910
/rpt_family="Alu"
14911. .15079
/rpt_family="L1"
15062. .15811
/rpt_family="L1"
15812. .16112
/rpt_family="Alu"
16113. .16529
/rpt_family="L1"
16530. .16941
/Note="similar to Homo sapiens EST AA446576 (NID:g2159241)
zw84d02.s1"
16541. .17068
/Note="similar to Homo sapiens EST BE966979
(NID:g11772953)"
16586. .17133
/Note="similar to Homo sapiens EST BE622588
(NID:g9893515)"
16610. .17149
/Note="similar to Homo sapiens EST A1744560 (NID:g5112848)
wg09g02.x1"
16743. .17248
/Note="similar to Homo sapiens EST N46675 (NID:g1187841)
yv50c07.r1"
16912. .17332
/Note="similar to Homo sapiens EST AA446703 (NID:g2159368)
zw84d02.r1"
16921. .17383
/Note="match to EST BG479148 (NID:g13411427)"
17135. .17383
/Note="match to EST BG773852 (NID:g14044151)"
17144. .17383
/Note="similar to Homo sapiens EST BG480641
(NID:g13412920)"
17383. .17822
/rpt_family="L1"
17826. .18132
/rpt_family="Alu"
17826. .17839
/Note="similar to Homo sapiens EST AA669082 (NID:g2630581)
ab92h09.s1"
18142. .18846
/rpt_family="ERV1"
18530. .18619
/Note="match to EST BF894500 (NID:g12285959)"
19083. .19168

Query Match 8.9%; Score 89; DB 9; Length 176547;
Best Local Similarity 68.0%; Pred. No. 2.le-15;
Matches 157; Conservative 0; Mismatches 65; Indels 9; Gaps 2;

Qy 771 CACGAGTGTAAAGACAGAGAGTCTGTGTGGGAATCTCCTTTGTGTGGATCATCAGGAG 830
Db 49359 CCCTGAGTGGAGAGAGAGGGTCTGTGCAGGAGAGTCTCCTGTGTGATCATGGGGTAA 49418

Qy 831 GTGAAGTCTTTGTATACGCTCATATCCAGCTTGTGTGATACCAATTCAGTGAAGCTGG 890
Db 49419 GTGATGCCTTTGGTCTGCATAAAGCCCTGGTGTGTGGAGCAGAGATTCGAGTGGGGCTGG 49478

Qy 891 AACAGCTGGCACTCTCTCAACAGGCCCTACCAAGACATCATGTTTTTTTTTTTTTTTCC 950
Db 49479 GACAAGCAGGTGCTGCTCAACAGGCCCTCAATACACC-----TTGTCCCCCTCCC 49530
```

---

```
Qy 951 ACCAAACCTGGACCTGAATGGGATGTGGACA-CACATAGAGTCCAGAGGA 1000
Db 49531 ACCATAGCTGACCTGGATGGGAATGTGACACCTCATAGACTCCAGAGGA 49581

RESULT 15
AC092634 194142 bp DNA linear PRI 01-MAR-2002
LOCUS Homo sapiens BAC clone RP11-340I6 from 7, complete sequence.
DEFINITION AC092634 AC055828
ACCESSION AC092634.3 GI:15809167
VERSION AC092634.3 GI:15809167
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE 1 (bases 1 to 194142)
AUTHORS Sulston,J.E. and Waterston,R.
TITLE Toward a complete human genome sequence
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)
MEDLINE 99063792
PUBMED 9847074
REFERENCE 2 (bases 1 to 194142)
AUTHORS Bielicki,D., Haakenson,W., Dignan,G., Elliott,G. and Holmes,A.
TITLE The sequence of Homo sapiens BAC clone RP11-340I6
JOURNAL Unpublished (2001)
REFERENCE 3 (bases 1 to 194142)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (19-JUL-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 4 (bases 1 to 194142)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (30-SEP-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 5 (bases 1 to 194142)
AUTHORS Waterston,R.H.
TITLE Direct Submission
JOURNAL Submitted (12-OCT-2001) Genome Sequencing Center, Washington
University School of Medicine, 4444 Forest Park Parkway, St. Louis,
MO 63108, USA
REFERENCE 6 (bases 1 to 194142)
AUTHORS Waterston,R.
TITLE Direct Submission
JOURNAL Submitted (01-MAR-2002) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Sep 30, 2001 this sequence version replaced gi:15624959.
COMMENT -----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: saplens@watson.wustl.edu
----- Summary Statistics -----
Center project name: H_NH0340I06
Drafting Center: WIBR
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.
```

### MAPPING INFORMATION:

**MAPPING INFORMATION:**  
The sequence of this clone was established as part of a mapping and sequencing collaboration between the NHGRI Chromosome 7 Mapping Project (Eric D. Green, Director), John D. McPherson in the Department of Genetics (Washington University), and the Washington University Genome Sequencing Center. For additional information about the map position of this sequence, see <http://www.nhgri.nih.gov/DIR/GRF/CHR7/>, send <mailto:egreen@nhgri.nih.gov>, or see <http://genome.wustl.edu/gsc>

**SOURCE INFORMATION:**

The RPI-11 human BAC library was made from the blood of one male donor, as described by Osogawa, K., Woon, P. Y., Zhao, B., Frengen, E., Tateno, M., Catanese, J. and de Jong, P. J. (1998). An improved approach for construction of bacterial artificial chromosome libraries. *Genomics* 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at <http://www.chori.org>

VECTOR: pBACs3.6

## NEIGHBORING SEQUENCE INFORMATION:

RELIGIONING SEQUENCE INFORMATION:  
The clone sequenced to the left is CTD-2526L21; the clone sequenced to the right is RP11-3N2. Actual start of this clone is at base position 1 of RP11-340I6; actual end is at base position 194112 of RP11-340I6;.

The sequence of AC055828 has been incorporated into AC092634.

## FEATURES

|   |  |
|---|--|
| location/Qualifiers                             |  |
| 1. .194142                                      |  |
| /organism="Homo sapiens"                        |  |
| /db_xref="taxon:9606"                           |  |
| /chromosome="7"                                 |  |
| /map="7"  |  |
| /clone="RP11-34016"                             |  |
| /clone.lib="RPC1-11"                            |  |
| 83. .392  |  |
| /rpt_family="Alu"                               |  |
| 399. .726                                       |  |
| /rpt_family="L1"                                |  |
| 735. .773                                       |  |
| /rpt_family="(TCCC)n"                           |  |
| 852. .1169                                      |  |
| /rpt_family="Alu"                               |  |
| 965. .1334                                      |  |
| /note="similar to EST BE147181 (NID:g8609905)"  |  |
| 1147. .1176                                     |  |
| /rpt_family="(CAAAA)n"                          |  |
| 1310. .1965                                     |  |
| /rpt_family="L1"                                |  |
| 2030. .2062                                     |  |
| /rpt_family="(CAPATA)n"                         |  |
| 2090. .2321                                     |  |
| /rpt_family="L1"                                |  |
| 2545. .2818                                     |  |
| /rpt_family="L1"                                |  |
| 2678. .2859                                     |  |
| /note="similar to EST AW992032 (NID:g8252112)"  |  |
| 2829. .2874                                     |  |
| /rpt_family="CT-rich"                           |  |
| 2902. .2955                                     |  |
| /rpt_family="AT-rich"                           |  |
| 2932. .2997                                     |  |
| /rpt_family="L1"                                |  |
| 2998. .3472                                     |  |
| /note="similar to EST AW992032 (NID:g8252112)"  |  |
| 3236. .3669                                     |  |
| /note="similar to EST AW753953 (NID:g7668885)"  |  |
| 3689. .3732                                     |  |
| /rpt_family="(TG)n"                             |  |
| 3844. .4141                                     |  |
| /rpt_family="Alu"                               |  |
| 3852. .4242                                     |  |
| /note="similar to EST AV760597 (NID:q10918445)" |  |

|               |               |   |
|---------------|---------------|---|
| misc_feature  | 4144. .4242   | /note="similar to EST AA609221 (NID:g2457649) afl2g08.sl" |
| misc_feature  | 4358. .4474   | /note="similar to EST AV760597 (NID:g10918445)"           |
| misc_feature  | 4258. .4466   | /note="similar to EST AA609221 (NID:g2457649) afl2g08.sl" |
| repeat_region | 4547. .4691   | /rpt_family="Alu"   |
| repeat_region | 4748. .5176   | /rpt_family="L1"  |
| repeat_region | 5361. .5394   | /rpt_family="(T)n"  |
| repeat_region | 5372. .5674   | /rpt_family="Alu"   |
| repeat_region | 5726. .6122   | /rpt_family="MaLR"  |
| repeat_region | 6125. .6294   | /rpt_family="Alu"   |
| repeat_region | 6302. .6820   | /rpt_family="(TA)n"                                       |
| repeat_region | 6825. .6951   | /rpt_family="Alu"   |
| repeat_region | 6982. .7292   | /rpt_family="Alu"   |
| repeat_region | 7266. .7315   | /rpt_family="(TAAA)n"                                     |
| repeat_region | 7670. .7964   | /rpt_family="Alu"   |
| repeat_region | 7967. .8030   | /rpt_family="MER1_type"                                   |
| repeat_region | 8278. .8356   | /rpt_family="(TCTA)n"                                     |
| repeat_region | 8332. .8394   | /rpt_family="ERVL"  |
| repeat_region | 8429. .8730   | /rpt_family="Alu"   |
| repeat_region | 9010. .9514   | /rpt_family="MaLR"  |
| repeat_region | 9515. .9809   | /rpt_family="Alu"   |
| repeat_region | 9811. .9832   | /rpt_family="AT_rich"                                     |
| repeat_region | 9816. .10107  | /rpt_family="Alu"   |
| repeat_region | 10108. .10144 | /rpt_family="MaLR"  |
| repeat_region | 10509. .10864 | /rpt_family="MaLR"  |
| repeat_region | 10871. .11071 | /rpt_family="L2"  |
| repeat_region | 11332. .11643 | /rpt_family="Alu"   |
| repeat_region | 12086. .12243 | /rpt_family="MER103"                                      |
| repeat_region | 12822. .13094 | /rpt_family="Alu"   |
| repeat_region | 13103. .13292 | /rpt_family="Alu"   |
| repeat_region | 14143. .14300 | /rpt_family="L2"  |
| repeat_region | 14301. .14603 |   |

|    | Query Match           | 8.7%   | Score 87.4;        | DB 9;          | Length 194142;    |
|----|-----------------------|--|--------------------|----------------|-------------------|
|    | Best Local Similarity | 67.5%;   | Pred. No. 6.5e-15; |                |                   |
|    | Matches 156;          | Conservative   | 0;                 | Mismatches 66; | Indels 9; Gaps 2; |
| QY | 771                   | CACGGAGTGTATAAAGAGAGAGATTCTGTGTCGGGAATCTCCTTTGGTGGATCATCAGGGAG | 830                |                |                   |
|    |                       |  |                    |                |                   |
| Db | 78694                 | CGTGAGTGGAGGAAGATAGGGTTCTGTGCAGGGAACCTCCCTTGGTGGATCATGGGGTAA   | 78753              |                |                   |
| QY | 831                   | GTGAAGTCTTTGTCAATAGCCTCATATCCAGCTTGTGTGATACCAATTCACAGTGAAGCTGG | 890                |                |                   |
|    |                       |  |                    |                |                   |
| Db | 78754                 | GTGATGCTGTGGTCTGACATAAAGCCCTGGTGTGTGGAGCAGAGATTCACATGGGGGTGG   | 78813              |                |                   |



**THIS PAGE BLANK (USPTO)**



GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 16:24:42 ; Search time 1812.02 Seconds  
(without alignments)  
16077.036 Million cell updates/sec

Title: US-09-708-724A-3\_COPY\_10000\_11000

Perfect score: 1001

Sequence: 1 caaaatttcagtttaggaaga.....gcagcacataatgtatcatg 1001

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: | 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%

Listing first 45 summaries

Database :

GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Match | Length | DB | ID       | Description        |
|------------|-------|-------|--------|----|----------|--------------------|
| 1          | 990   | 98.9  | 151049 | 9  | AC018558 | AC018558 Homo sapi |
| 2          | 990   | 98.9  | 191768 | 9  | AC092357 | AC092357 Homo sapi |
| 3          | 631.6 | 63.1  | 144577 | 9  | AC023824 | AC023824 Homo sapi |
| 4          | 631.6 | 63.1  | 201061 | 9  | AC007353 | AC007353 Homo sapi |
| 5          | 601   | 60.0  | 191014 | 9  | AC113268 | AC113268 Homo sapi |
| 6          | 578.8 | 57.8  | 167133 | 2  | AC026186 | AC026186 Homo sapi |
| 7          | 555   | 55.4  | 167133 | 2  | AC026186 | AC026186 Homo sapi |
| 8          | 371.6 | 37.1  | 77060  | 2  | AC025998 | AC025998 Homo sapi |
| 9          | 365.2 | 36.5  | 77060  | 2  | AC025998 | AC025998 Homo sapi |
| 10         | 234.4 | 23.4  | 197915 | 9  | AC005725 | AC005725 Homo sapi |
| 11         | 101.4 | 10.1  | 193259 | 9  | AC025281 | AC025281 Homo sapi |
| 12         | 97.6  | 9.8   | 214269 | 9  | AC016716 | AC016716 Homo sapi |
| 13         | 96.2  | 9.6   | 112687 | 9  | AC008632 | AC008632 Homo sapi |
| 14         | 96.2  | 9.6   | 195470 | 2  | AC022259 | AC022259 Homo sapi |
| 15         | 95.2  | 9.5   | 147330 | 9  | AC093514 | AC093514 Homo sapi |
| 16         | 95.2  | 9.5   | 179636 | 9  | AC026462 | AC026462 Homo sapi |
| 17         | 94.6  | 9.5   | 167227 | 9  | AC020602 | AC020602 Homo sapi |
| 18         | 92.4  | 9.2   | 127025 | 9  | AC018880 | AC018880 Homo sapi |
| 19         | 92.4  | 9.2   | 203407 | 2  | AC006174 | AC006174 Homo sapi |
| 20         | 90.4  | 9.0   | 173719 | 2  | AC108826 | AC108826 Homo sapi |
| 21         | 90.4  | 9.0   | 176223 | 9  | AC009021 | AC009021 Homo sapi |
| 22         | 90.4  | 9.0   | 192609 | 9  | AC009150 | AC009150 Homo sapi |
| 23         | 90.4  | 9.0   | 208884 | 2  | AC019092 | AC019092 Homo sapi |
| 24         | 90.2  | 9.0   | 31994  | 9  | AC011528 | AC011528 Homo sapi |
| 25         | 90.2  | 9.0   | 167791 | 9  | AC108075 | AC108075 Homo sapi |
| 26         | 89.6  | 9.0   | 154158 | 2  | AL136527 | AL136527 Homo sapi |
| 27         | 89.6  | 9.0   | 191824 | 9  | AL161935 | AL161935 Human DNA |
| 28         | 89.4  | 8.9   | 114800 | 9  | AC123567 | AC123567 Homo sapi |
| 29         | 89.4  | 8.9   | 143618 | 2  | AC026855 | AC026855 Homo sapi |
| 30         | 89.4  | 8.9   | 162808 | 2  | AC021340 | AC021340 Homo sapi |
| 31         | 89.4  | 8.9   | 168872 | 9  | AC073325 | AC073325 Homo sapi |
| 32         | 89.4  | 8.9   | 199451 | 2  | AC027045 | AC027045 Homo sapi |
| 33         | 89.4  | 8.9   | 210719 | 2  | AC087646 | AC087646 Homo sapi |
| 34         | 89.4  | 8.9   | 213359 | 2  | AC069006 | AC069006 Homo sapi |
| 35         | 89.4  | 8.9   | 214738 | 30 | AC023339 | AC023339 Homo sapi |
| 36         | 89.2  | 8.9   | 100226 | 9  | AL583802 | AL583802 Human DNA |
| 37         | 89.2  | 8.9   | 172420 | 9  | AC022182 | AC022182 Homo sapi |
| 38         | 89.2  | 8.9   | 180541 | 2  | AC016521 | AC016521 Homo sapi |
| 39         | 88.4  | 8.8   | 135940 | 2  | AC024294 | AC024294 Homo sapi |
| 40         | 88.4  | 8.8   | 158095 | 2  | AC016224 | AC016224 Homo sapi |
| 41         | 88.4  | 8.8   | 160615 | 9  | AC112236 | AC112236 Homo sapi |
| 42         | 88.4  | 8.8   | 163974 | 9  | AB011399 | AB011399 Homo sapi |
| 43         | 88.4  | 8.8   | 169184 | 9  | AC022398 | AC022398 Homo sapi |
| 44         | 88.4  | 8.8   | 199385 | 2  | AC019049 | AC019049 Homo sapi |
| 45         | 88.4  | 8.8   | 201088 | 9  | AL589733 | AL589733 Human DNA |

ALIGNMENTS

RESULT 1

AC018558

LOCUS

DEFINITION

AC018558

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AC018558 151049 bp DNA linear PRI 02-NOV-2001  
Homo sapiens chromosome 16 clone RP11-80F22, complete sequence.

AC018558

AC018558.5 GI:16596530

HTG.

Homo sapiens.

Homo sapiens

Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 151049)

DOE Joint Genome Institute.

Sequencing of Human Chromosome 16

Unpublished

2 (bases 1 to 151049)  
 DOE Joint Genome Institute.  
 Direct Submission  
 Submitted (14-DEC-1999) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

3 (bases 1 to 151049)  
 DOE Joint Genome Institute.  
 Direct Submission  
 Submitted (02-NOV-2001) Production Sequencing Facility, DOE Joint  
 Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 On Nov 2, 2001 this sequence version replaced gi:9795566.

Sequence Quality Assessment:  
 This entry has been annotated with sequence quality  
 estimates computed by the Phrap assembly program.  
 All manually edited bases have been reduced to quality zero.  
 Quality levels above 40 are expected to have less than  
 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the  
 GenBank flat file format but are available as part  
 of this entry's ASN.1 file.

-----  
 Sequence Quality Assessment:  
 This entry has been annotated with sequence quality  
 estimates computed by the Phrap assembly program.  
 All manually edited bases have been reduced to quality zero.  
 Quality levels above 40 are expected to have less than  
 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the  
 GenBank flat file format but are available as part  
 of this entry's ASN.1 file.

-----

FEATURES  
 Location/Qualifiers  
 source 1..151049  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="16"  
 /clone="RP11-80F22"

BASE COUNT 46827 a 30783 c 30749 g 42690 t

ORIGIN

Query Match 98.9%; Score 990; DB 9; Length 151049;  
 Best Local Similarity 99.9%; Pred. No. 6.6e-231;  
 Matches 1001; Conservative 0; Mismatches 0; Indels 1; Gaps 1

QY 1 CAAATTTCAGTTAGGAAGAATAAGTGAAGAGATCTATTGTACTTGGTACTACAGTTA 60  
 DB 80220 CAAATTTCAGTTAGGAAGAATAAGTGAAGAGATCTATTGTACTTGGTACTACAGTTA 80279

QY 61 ATCTATTGTGTTCTTGACTAATACAGTAGATTTCGAGTGTCTTCACAAACAAACATGAT 120  
 DB 80280 ATGTATTGTGTTCTTGACTAATACAGTAGATTTCGAGTGTCTTCACAAACAAACATGAT 80339

QY 121 GGGTATGTAGGTAATGCATATGCAACATAGCTTGGGTTAACCAATTCACAATATGTGTG 180  
 DB 80340 GGGTATGTAGGTAATGCATATGCAACATAGCTTGGGTTAACCAATTCACAATATGTGTG 80399

QY 181 TATTTCAAACACGTACCATAAATGCAGACAATTTTGTCTCAGTTACAAATCAAAAAGTTTT 240  
 DB 80400 TATTTCAAACACGTACCATAAATGCAGACAATTTTGTCTCAGTTACAAATCAAAAAGTTTT 80459

QY 241 TAAATGAGGACCTTAGGTGGGTCTCTAATCAATCTTAAGTGTCTTCCATGAAGAGG 300  
 DB 80460 TAAATGAGGACCTTAGGTGGGTCTCTAATCAATCTTAAGTGTCTTCCATGAAGAGG 80519

QY 301 AAATAAGGATACAAATGTGCACACAGAGAGAAATGGCCACATGAGGACACAATGAGAATG 360  
 DB 80520 AAATAAGGATACAAATGTGCACACAGAGAGAAATGGCCACATGAGGACACAATGAGAATG 80579

QY 361 TGCTACTTACAGCCTTAGAGAGAGGGCTCCGAGAAACACACACCCCTACCCACACCTTGA 420  
 DB 80580 TGCTACTTACAGCCTTAGAGAGAGGGCTCCGAGAAACACACACCCCTACCCACACCTTGA 80639

QY 421 TGTGGACTTCATCCCTGTAGAGAGAAAGTCTCTCCACCCCTCCTTCATCAGGTGGAAGCCCTTG 480



FEATURES  
source

-----  
Location/Qualifiers  
1. 144577  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-244B22"  
BASE COUNT 39644 a 31376 c 31688 g 41869 t  
ORIGIN  
  
Query Match 63.1%; Score 631.6; DB 9; Length 144577;  
Best Local Similarity 80.6%; Pred. No. 1.3e-143;  
Matches 830; Conservative 0; Mismatches 154; Indels 46; Gaps 6;  
  
QY 10 AGTTAGGAAGATAAGTCAAGAGACTATTTGTTGCTGCTACAGTTAATGTTATCT 69  
Db 29512 AGACAGGAGAAATAGGACATCAGATCTATGTCATTTGGTGACTACAGTTAATGTTATCT 29453  
QY 70 GTTCTTGA-----CTAATACAGTAGATTTCGAGTCTTCTCAACAAAAACATGATGG 123  
Db 29452 GTTCTTGAACATGCTAACACCTGTAGATTTTGAGTCTCTCAACACAGAAAAATGATGG 29393  
QY 124 TATGTCAGTAAATGATATGAACTAGCTTGGTTAAACATCCCAATATGTTGTTAT 183  
Db 29392 TATGTCAGTAAATGATATGAACTAGCTTGGTTAAACATCCCAATATGTTGTTAT 29333  
QY 184 TTCAAAACAGTACCAATAATGACAGCAATTTT-----GTGTCAGTTACAATCAAAAAAG 237  
Db 29332 TTAAAAACAGTCCCAAAATGTAGACAATTTTATTAGTTACTAGTTACAATAAAAAATG 29273  
QY 238 TTTTAAATCAGGACCTTAGGTGGTCTTAATCCCAATCTAAGTGTATGTTCCATGAAAG 297  
Db 29272 TTTTAAATCAGTACATTAGGTGGTGGCCCTGATCCCAATCTAAGTGTATGTTCCATGAAAG 29213  
QY 298 AGCAATATAGGATACAAATGTGCACACAGAGAAATGCCACATGAGACACATGACA 357  
Db 29212 AGGAGATAAGGATACAAATGTGCACACAGAGAAATGCCACATGAGACACATGACA 29153  
QY 358 ATGTGCTACTTACAAGCTTAGGAGAGCGCTCCAGAGAAAACACACCTTACCACACCT. 417  
Db 29152 AGGTGCCACTTACACACCTTAGGAGAGAGCGCTTGGGGCACACACCTTGGCCACCT 29093  
QY 418 TGATGTTGA-----CTTCATCTGTAGACAGAAAGTCTCCACCTTCCACCC 456  
Db 29092 TGACCTGGACTTCACTCCCCAGATAAAGTCTCTTCACTCCAGATAAAGTCTTCTATCC 29033  
QY 457 TCCTTCATCAGTGGAGCGCTTTGATTTCTCAATATTTCTCAATGCTGGAAGGTACAAA 516  
Db 29032 TCCTTCATCATATGGAACCTTCTGACCCCTGAATACCTCAATGCTGTAATGTACCAAG 28973  
QY 517 GTGAAGAGACAGACAGACCTCAGGCTGAAAGTTTAAAGAGAAATTAACATCTTCCATTG 576  
Db 28972 GTGAAGGACAGACAGACCTCAGGCTGAAAGTTTAAAGAGAAATTAACATCTTCCATTG 28913  
QY 577 CTGTGCTCTATCCCTTACACACCTTATCCAGTCTTTTATGCTCTTTTGTGTTTTCGTTG 636  
Db 28912 CCCTGTCCCATCCCAACACACACCTGTCCAGCCCTTATGCTCTTTTGTATTCCTCCCTG 28853  
QY 637 TCCTGGGTATAGTGTAGTTGTAATCTGTGTTTACATACAGGATACATAAACAAGG 696  
Db 28852 TCCTGGATATAGGTTAACTTATTAATCTGTGTTTATGTATGATTAACATAAACAAGG 28793  
QY 697 TAAACAATAAATAAACAACAGACAGCAAACTCAACTAATAGTGTGTTGGCATGTTGACA 756  
Db 28792 TAAACAATAAATAAATAAACAACAGCAAACTCAACTAATAGTGTGTTGGCATGTTGACA 28733  
QY 757 GTGAAGACAGGAGTACATAAATAATGGAGTGGAACT--TTTGTAGCTAAATCAATGTCC 815  
Db 28732 GTGAAGGACAGGAGATCACATACAAATGGAGTGGAGTCTTGGAGCTCAGTCCACATCC 28673  
QY 816 TGTGTGTTCTGTTTCTACATCAGACTCTATAGTGGCAATGTTCAGTGTAGTGTGTT 875  
Db 28672 TGT-----TGTGTTTCTACATCAAGATCTATAGTACAGTATTTCAGGTTTGTGTT 28621

QY 876 TTATCTTGCCTGCTCAGTAAGTGCAGAGAGATTTTCTAACTGGGTGAGAACAGG 935  
Db 28620 TTGTCTTGCCTGCCCACTAAGTGCAGAGGGGATTTCTAATCAGGTGAGGACTCG 28561  
QY 936 TA----GAGAAGTGAAGTGAGACAAACTTCCTGCCAFTTCCCAAGTGGCAGCACATA 991  
Db 28560 TAACTTGCACTTTGCAAGTGCAGACAACTTCTCTTGCCATTTGCCACAGTGGCATGCGACA 28501  
QY 992 ATGTATCATG 1001  
Db 28500 ATTTATCATG 28491  
  
RESULT 4  
AC007353/c  
LOCUS AC007353 201061 bp DNA linear PRI 01-JUN-2001  
DEFINITION Homo sapiens chromosome 16 clone RP11-488I20, complete sequence.  
ACCESSION AC007353  
VERSION AC007353.4 GI:14277247  
KEYWORDS HTG  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 201061)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 16  
JOURNAL Unpublished  
AUTHORS  
REFERENCE 2 (bases 1 to 201061)  
JOURNAL Unpublished  
AUTHORS Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,  
Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,  
Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,  
Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,  
McMurry,K., Han,C. and Deaven,L.  
TITLE Direct Submision  
JOURNAL Submitted (22-APR-1999) Center for Human Genome Studies, DOE Joint  
Genome Institute, Los Alamos National Laboratory, MS M888, Los  
Alamos, NM 87545, USA  
REFERENCE 3 (bases 1 to 201061)  
JOURNAL Direct Submision  
AUTHORS DOE Joint Genome Institute.  
TITLE Submitted (01-JUN-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT On Jun 1, 2001 this sequence version replaced gi:979555.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.  
-----  
Location/Qualifiers  
1. 201061  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-488I20"  
BASE COUNT 57062 a 41994 c 41012 g 60993 t  
ORIGIN

FEATURES  
source

Query Match 63.1%; Score 631.6; DB 9; Length 201061;  
Best Local Similarity 80.6%; Pred. No. 1.3e-143;  
Matches 830; Conservative 0; Mismatches 154; Indels 46; Gaps 6;  
  
QY 876 TTATCTTGCCTGCTCAGTAAGTGCAGAGAGATTTTCTAACTGGGTGAGAACAGG 935  
Db 28620 TTGTCTTGCCTGCCCACTAAGTGCAGAGGGGATTTCTAATCAGGTGAGGACTCG 28561  
QY 936 TA----GAGAAGTGAAGTGAGACAAACTTCCTGCCAFTTCCCAAGTGGCAGCACATA 991  
Db 28560 TAACTTGCACTTTGCAAGTGCAGACAACTTCTCTTGCCATTTGCCACAGTGGCATGCGACA 28501  
QY 992 ATGTATCATG 1001  
Db 28500 ATTTATCATG 28491  
  
RESULT 4  
AC007353/c  
LOCUS AC007353 201061 bp DNA linear PRI 01-JUN-2001  
DEFINITION Homo sapiens chromosome 16 clone RP11-488I20, complete sequence.  
ACCESSION AC007353  
VERSION AC007353.4 GI:14277247  
KEYWORDS HTG  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 201061)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 16  
JOURNAL Unpublished  
AUTHORS  
REFERENCE 2 (bases 1 to 201061)  
JOURNAL Unpublished  
AUTHORS Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,  
Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,  
Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,  
Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R.,  
McMurry,K., Han,C. and Deaven,L.  
TITLE Direct Submision  
JOURNAL Submitted (22-APR-1999) Center for Human Genome Studies, DOE Joint  
Genome Institute, Los Alamos National Laboratory, MS M888, Los  
Alamos, NM 87545, USA  
REFERENCE 3 (bases 1 to 201061)  
JOURNAL Direct Submision  
AUTHORS DOE Joint Genome Institute.  
TITLE Submitted (01-JUN-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT On Jun 1, 2001 this sequence version replaced gi:979555.  
-----Genome Center  
Center: Joint Genome Institute  
Center Code: JGI  
Web site: http://www.jgi.doe.gov  
-----  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.  
-----  
Location/Qualifiers  
1. 201061  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-488I20"  
BASE COUNT 57062 a 41994 c 41012 g 60993 t  
ORIGIN





|              |         |                                |           |         |               |
|--------------|---------|--------------------------------|-----------|---------|---------------|
| *            | 27497   | 27596:                         | gap of    | unknown | length        |
| *            | 27597   | 30050:                         | contig of | 2454    | bp in length  |
| *            | 30150   | 32351:                         | gap of    | unknown | length        |
| *            | 30151   | 32351:                         | contig of | 2201    | bp in length  |
| *            | 32352   | 32451:                         | gap of    | unknown | length        |
| *            | 32452   | 34424:                         | contig of | 1973    | bp in length  |
| *            | 34425   | 34524:                         | gap of    | unknown | length        |
| *            | 34525   | 36830:                         | contig of | 2306    | bp in length  |
| *            | 36831   | 36930:                         | gap of    | unknown | length        |
| *            | 36931   | 39686:                         | contig of | 2756    | bp in length  |
| *            | 39687   | 39786:                         | gap of    | unknown | length        |
| *            | 39787   | 42513:                         | contig of | 2726    | bp in length  |
| *            | 42513   | 42612:                         | gap of    | unknown | length        |
| *            | 42613   | 44556:                         | contig of | 1944    | bp in length  |
| *            | 44557   | 44656:                         | gap of    | unknown | length        |
| *            | 44657   | 48395:                         | contig of | 3739    | bp in length  |
| *            | 48396   | 48495:                         | gap of    | unknown | length        |
| *            | 48496   | 50933:                         | contig of | 2440    | bp in length  |
| *            | 50936   | 51035:                         | gap of    | unknown | length        |
| *            | 51036   | 54435:                         | contig of | 3400    | bp in length  |
| *            | 54436   | 54535:                         | gap of    | unknown | length        |
| *            | 54536   | 59749:                         | contig of | 5214    | bp in length  |
| *            | 59750   | 59849:                         | gap of    | unknown | length        |
| *            | 59850   | 63021:                         | contig of | 3172    | bp in length  |
| *            | 63022   | 63121:                         | gap of    | unknown | length        |
| *            | 63122   | 66858:                         | contig of | 3737    | bp in length  |
| *            | 66859   | 66958:                         | gap of    | unknown | length        |
| *            | 66959   | 71348:                         | contig of | 4390    | bp in length  |
| *            | 71349   | 71448:                         | gap of    | unknown | length        |
| *            | 71449   | 75783:                         | contig of | 4335    | bp in length  |
| *            | 75784   | 75883:                         | gap of    | unknown | length        |
| *            | 75884   | 79500:                         | contig of | 3617    | bp in length  |
| *            | 79501   | 79600:                         | gap of    | unknown | length        |
| *            | 79601   | 83515:                         | contig of | 3916    | bp in length  |
| *            | 83517   | 83616:                         | gap of    | unknown | length        |
| *            | 83617   | 88942:                         | contig of | 5326    | bp in length  |
| *            | 88943   | 89042:                         | gap of    | unknown | length        |
| *            | 89043   | 95244:                         | contig of | 6202    | bp in length  |
| *            | 95245   | 95344:                         | gap of    | unknown | length        |
| *            | 95345   | 99266:                         | contig of | 3922    | bp in length  |
| *            | 99267   | 99366:                         | gap of    | unknown | length        |
| *            | 99367   | 104303:                        | contig of | 4937    | bp in length  |
| *            | 104304  | 104403:                        | gap of    | unknown | length        |
| *            | 104404  | 110029:                        | contig of | 5626    | bp in length  |
| *            | 110030  | 110129:                        | gap of    | unknown | length        |
| *            | 110130  | 117586:                        | contig of | 7457    | bp in length  |
| *            | 117587  | 117686:                        | gap of    | unknown | length        |
| *            | 117687  | 125138:                        | contig of | 7452    | bp in length  |
| *            | 125139  | 1325238:                       | gap of    | unknown | length        |
| *            | 1325239 | 1329290:                       | contig of | 7052    | bp in length  |
| *            | 1329291 | 1332990:                       | gap of    | unknown | length        |
| *            | 1332991 | 1427118:                       | contig of | 10328   | bp in length  |
| *            | 142719  | 142818:                        | gap of    | unknown | length        |
| *            | 142819  | 154592:                        | contig of | 11774   | bp in length  |
| *            | 154593  | 1546592:                       | gap of    | unknown | length        |
| *            | 154693  | 167133:                        | contig of | 12441   | bp in length. |
| FEATURES     |         |                                |           |         |               |
| source       |         | Location/Qualifiers            |           |         |               |
|              |         | 1..167133                      |           |         |               |
|              |         | /organism="Homo sapiens"       |           |         |               |
|              |         | /db_xref="taxon:9606"          |           |         |               |
|              |         | /chromosome="3"                |           |         |               |
|              |         | /map="3p"                      |           |         |               |
|              |         | /clone="rp11-303p8"            |           |         |               |
| misc_feature |         | 1..2054                        |           |         |               |
|              |         | /note="assembly_name:Contig17" |           |         |               |
| misc_feature |         | 2155..3414                     |           |         |               |
|              |         | /note="assembly_name:Contig18" |           |         |               |
| misc_feature |         | 3515..5281                     |           |         |               |
|              |         | /note="assembly_name:Contig19" |           |         |               |
| misc_feature |         | 5392..6678                     |           |         |               |
|              |         | /note="assembly_name:Contig21" |           |         |               |
| misc_feature |         | 6779..8533                     |           |         |               |
|              |         | /note="assembly_name:Contig22" |           |         |               |



QY 541 GGTGAAAGTTTAAAGAGAAATACATCTTTCCATCTGCTGTGCTCCATCCCTACACAC 600  
|||||  
Db 137061 GGTGAAAGTTTAAAGAGAAATACACCTCCCATTCGCCCTGCTCCATCCCAACACACAC 137002  
QY 601 CTAATCCAGTCCTTATGTCCTTTTGTGTTTTCGTCCTCGGTATAGTGTAGTTGTTAA 660  
|||||  
Db 137001 CTGTGCCAGCCTTTATGTTGTTTGTGCTTCCCTCTCCCTGGGATAGTGTAACTGTTG 136942  
QY 661 ATCTGTGTTTACATACAGATACATATAAACAAGGTAACAATAAATAAACAAGACACA 720  
|||||  
Db 136941 ATTGTGTTTACATAGATACATGTAACAAAGGTAACAATAAATAAACAAGACACA 136882  
QY 721 GCAAACTCAACTAATAGTGTGTTGGCATGTCACAGTGAACAGACAGGAGTCAACATAA 780  
|||||  
Db 136881 GCACAACTCAATAATAGTGTGTTGGCGGGTGACAGTGAAGGAAAG-GCCACATAC 136823  
QY 781 ATGGAGGTGGAC-TTTTGAGCTAAATCAATGCTGTTGTTCTGTGTTCTTACAT 839  
|||||  
Db 136822 AATGGAGATGGAGCTTTTGAACATAATCAATGCTGTTGTT-----TGTTCCTACAT 136772  
QY 840 CAGACTCTATAGTGGCAATGTCAGTGTAGTGTGTTTATCCTGCTGCTCAGTAAAGTG 899  
|||||  
Db 136771 CAGAATCTATAGTACAAATGTCAGTGTAGTGTGTTTGTCTTATCTCCCATTAAGTG 136712  
QY 900 CCAGAGGAGATTTTCTAAACTGGGTGAGGAACAGGTAGAGTGTAAAGTGAAGACAAAC 959  
|||||  
Db 136711 CCAGAGGGGATTTACTTAAACTGGGTGAGGAACAAAGTCAAGGAGT---GTGAGACAAAC 136656  
QY 960 TTCCCTGCCATTTGCCAAGTGGCAGCAGCATATATGATCATG 1001  
|||||  
Db 136655 TTCCCTGCCATTTGCCACAGTGGCAGGAGCAATTCATCATG 136614

RESULT 7  
AC026186  
LOCUS 167133 bp DNA linear HTG 29-MAY-2000  
DEFINITION Homo sapiens chromosome 3 clone RP11-303P8 map 3p, WORKING DRAFT  
SEQUENCE, 44 unordered pieces.  
AC026186  
AC026186.2 GI:8101183  
VERSION HTG: HTGS\_PHASE1; HTGS\_DRAFT.  
KEYWORDS Homo sapiens.  
SOURCE Homo sapiens  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 167133)  
AUTHORS Bao,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,  
Dong,W., Fan,H., Feng,X., Guan,Q., Gu,X., Guo,D., He,L., Hu,S.,  
Huang,F., Jin,Y., Kang,N., Li,C., Li,C., Li,G., Li,J., Li,L.,  
Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,X., Liu,Y., Li,W., Li,Y.,  
Luo,J., Niu,Y., Qi,Q., Qi,X., Song,S., Sun,M., Sun,W., Sun,Y.,  
Tao,R., Wang,H., Wang,J., Wang,J., Wang,L., Wang,L., Wang,R.,  
Wang,X., Wang,X., Wang,Y., Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y.,  
Yan,C., Yang,X., Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H.,  
Zhang,L., Zhang,M., Zhang,X., Zhang,X., Zhang,Y., Zhang,Y.,  
Zhang,Z., Zhu,B., Yu,J. and Yang,H.  
Chromosome 3p genomic sequence  
TITLE  
JOURNAL  
REFERENCE 2 (bases 1 to 167133)  
AUTHORS Chen,J., Hu,S., Dong,W., Wang,J., Zhang,Y., Zhang,H., Liu,B.,  
Bao,W., Sun,Y., Wu,Q., Wang,H., Yang,X., Cheng,C., Wang,Y., Niu,Y.,  
Qi,X., Li,T., Zhang,H., Liu,N., Wu,D., Yu,B., Fan,H., Liu,Y.,  
Li,G., Li,C., Bao,Q., Bao,J., Wang,X., Song,L., Zhang,L., Guo,D.,  
Huang,F., Zhang,S., Li,J., Bian,X., Zhang,M., Li,L., Feng,X., Yu,J.  
and Yang,H.  
Direct Submission  
JOURNAL  
TITLE  
JOURNAL Submitted (21-MAR-2000) Human Genomic Center, Institute of  
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing  
100101, P.R.China  
COMMENT On May 29, 2000 this sequence version replaced gi:7272003.  
-----Genome Center  
Center:Beijing Center

Center code:Beijing  
Website:http://hg.c.igtp.ac.cn  
http://www.genomics.org.cn  
Contact:hgc@igtp.ac.cn  
----- Project Information  
Center project name:1# project  
Center clone name: RP11-303P8  
----- Summary Statistics  
Sequencing vector: pUC18; 100% of reads  
Chemistry: Dye-terminator: ET 5% of reads  
Chemistry: Dye-terminator Big Dye; 45% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 144681 bases at least Q40  
Consensus quality: 152516 bases at least Q30  
Consensus quality: 159163 bases at least Q20  
Insert size: 121837; sum-of-contigs  
Quality coverage: 4.43x in Q20 bases;sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 44 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 2054: contig of 2054 bp in length  
\* 2055 2154: gap of unknown length  
\* 2155 3414: contig of 1260 bp in length  
\* 3415 5281: contig of 1767 bp in length  
\* 5282 5382: gap of unknown length  
\* 5382 6678: contig of 1297 bp in length  
\* 6679 6779: gap of unknown length  
\* 6779 8533: contig of 1755 bp in length  
\* 8534 8633: gap of unknown length  
\* 8634 9989: contig of 1356 bp in length  
\* 9990 10089: gap of unknown length  
\* 10090 11298: contig of 1209 bp in length  
\* 11299 11398: gap of unknown length  
\* 11399 12931: contig of 1333 bp in length  
\* 12932 13031: gap of unknown length  
\* 13032 14757: contig of 1725 bp in length  
\* 14757 14856: gap of unknown length  
\* 14857 16838: contig of 1982 bp in length  
\* 16839 16938: gap of unknown length  
\* 16939 18630: contig of 1692 bp in length  
\* 18631 18730: gap of unknown length  
\* 18731 20330: contig of 1600 bp in length  
\* 20331 20430: gap of unknown length  
\* 20431 21642: contig of 1212 bp in length  
\* 21643 21742: gap of unknown length  
\* 21743 23363: contig of 1621 bp in length  
\* 23364 23463: gap of unknown length  
\* 23464 25148: contig of 1685 bp in length  
\* 25149 25248: gap of unknown length  
\* 25249 27496: contig of 2248 bp in length  
\* 27497 27596: gap of unknown length  
\* 27597 30050: contig of 2454 bp in length  
\* 30051 30150: gap of unknown length  
\* 30151 32351: contig of 2201 bp in length  
\* 32352 32451: gap of unknown length  
\* 32452 34424: contig of 1973 bp in length  
\* 34425 34524: gap of unknown length  
\* 34525 36830: contig of 2306 bp in length  
\* 36831 36930: gap of unknown length  
\* 36931 39686: contig of 2756 bp in length  
\* 39687 39786: gap of unknown length  
\* 39787 42612: contig of 2726 bp in length  
\* 42613 44556: contig of 1944 bp in length  
\* 44557 44656: gap of unknown length  
\* 44657 48395: contig of 3739 bp in length





- \* overlap relationships among clones to be deduced.
- \* However, it should not be assumed that this clone
- \* will be sequenced to completion. In the event that
- \* the record is updated, the accession number will
- \* be preserved.

|   |       |               |                            |                            |
|---|-------|---------------|----------------------------|----------------------------|
| * | 1     | be preserved. | 773:                       | contig of 773 bp in length |
| * | 774   | 873:          | gap of 100 bp              |                            |
| * | 874   | 1684:         | contig of 811 bp in length |                            |
| * | 1685  | 1784:         | gap of 100 bp              |                            |
| * | 1785  | 2572:         | contig of 788 bp in length |                            |
| * | 2573  | 2672:         | gap of 100 bp              |                            |
| * | 2673  | 3467:         | contig of 795 bp in length |                            |
| * | 3468  | 3567:         | gap of 100 bp              |                            |
| * | 3568  | 4354:         | contig of 787 bp in length |                            |
| * | 4355  | 4454:         | gap of 100 bp              |                            |
| * | 4455  | 5254:         | contig of 800 bp in length |                            |
| * | 5255  | 5354:         | gap of 100 bp              |                            |
| * | 5355  | 6140:         | contig of 786 bp in length |                            |
| * | 6141  | 6240:         | gap of 100 bp              |                            |
| * | 6241  | 7045:         | contig of 805 bp in length |                            |
| * | 7046  | 7145:         | gap of 100 bp              |                            |
| * | 7146  | 7932:         | contig of 787 bp in length |                            |
| * | 7933  | 8032:         | gap of 100 bp              |                            |
| * | 8033  | 8851:         | contig of 819 bp in length |                            |
| * | 8852  | 8951:         | gap of 100 bp              |                            |
| * | 8952  | 9754:         | contig of 803 bp in length |                            |
| * | 9755  | 9854:         | gap of 100 bp              |                            |
| * | 9855  | 10640:        | contig of 786 bp in length |                            |
| * | 10641 | 10740:        | gap of 100 bp              |                            |
| * | 10741 | 11547:        | contig of 807 bp in length |                            |
| * | 11548 | 11647:        | gap of 100 bp              |                            |
| * | 11648 | 12448:        | contig of 801 bp in length |                            |
| * | 12449 | 12548:        | gap of 100 bp              |                            |
| * | 12549 | 13346:        | contig of 798 bp in length |                            |
| * | 13347 | 13446:        | gap of 100 bp              |                            |
| * | 13447 | 14234:        | contig of 788 bp in length |                            |
| * | 14235 | 14334:        | gap of 100 bp              |                            |
| * | 14335 | 15106:        | contig of 772 bp in length |                            |
| * | 15107 | 15206:        | gap of 100 bp              |                            |
| * | 15207 | 16013:        | contig of 807 bp in length |                            |
| * | 16014 | 16113:        | gap of 100 bp              |                            |
| * | 16114 | 16924:        | contig of 811 bp in length |                            |
| * | 16925 | 17024:        | gap of 100 bp              |                            |
| * | 17025 | 17798:        | contig of 774 bp in length |                            |
| * | 17799 | 17898:        | gap of 100 bp              |                            |
| * | 17899 | 18706:        | contig of 808 bp in length |                            |
| * | 18707 | 18806:        | gap of 100 bp              |                            |
| * | 18807 | 19616:        | contig of 810 bp in length |                            |
| * | 19617 | 19716:        | gap of 100 bp              |                            |
| * | 19717 | 20523:        | contig of 807 bp in length |                            |
| * | 20524 | 20623:        | gap of 100 bp              |                            |
| * | 20624 | 21400:        | contig of 777 bp in length |                            |
| * | 21401 | 21500:        | gap of 100 bp              |                            |
| * | 21501 | 22285:        | contig of 785 bp in length |                            |
| * | 22286 | 22385:        | gap of 100 bp              |                            |
| * | 22386 | 23185:        | contig of 800 bp in length |                            |
| * | 23186 | 23285:        | gap of 100 bp              |                            |
| * | 23286 | 24103:        | contig of 818 bp in length |                            |
| * | 24104 | 24203:        | gap of 100 bp              |                            |
| * | 24204 | 25012:        | contig of 809 bp in length |                            |
| * | 25013 | 25112:        | gap of 100 bp              |                            |
| * | 25113 | 25922:        | contig of 810 bp in length |                            |
| * | 25923 | 26022:        | gap of 100 bp              |                            |
| * | 26023 | 26839:        | contig of 817 bp in length |                            |
| * | 26840 | 26939:        | gap of 100 bp              |                            |
| * | 26940 | 27765:        | contig of 826 bp in length |                            |
| * | 27766 | 27865:        | gap of 100 bp              |                            |
| * | 27866 | 28662:        | contig of 797 bp in length |                            |
| * | 28663 | 28762:        | gap of 100 bp              |                            |
| * | 28763 | 29574:        | contig of 812 bp in length |                            |
| * | 29575 | 29674:        | gap of 100 bp              |                            |
| * | 29675 | 30453:        | contig of 779 bp in length |                            |
| * | 30454 | 30553:        | gap of 100 bp              |                            |

[illegible]

**Qy**   909 ATTTTCTAAACGGGTAGGGAACAGGTAGAGAAGTGTAAGTGAGACAACATCCCTGCC 968  
           ||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||  
**Dd** 102343 ATTATTCTAATAATGGGTGAGGAACAGGTAGAGAAGTGTAAGTGAGACAACATTCCTGCC 102402

Qy 969 ATTTGCCAAAGTGGCAGCACATAATGTATCATG 1001  
||||||| ||||||| || ||| |||||||  
Db 102403 ATTTGCCACAGTGGCAGGGCAGAAATTCATCATG 102435

|            |   |          |     |        |                 |
|------------|---|----------|-----|--------|-----------------|
| RESULT 8   |   |          |     |        |                 |
| AC025998/c |   |          |     |        |                 |
| LOCUS      | AC025998  | 77060 bp | DNA | linear | HTG 10-MAR-2000 |
| DEFINITION | Homo sapiens chromosome 10 clone RP11-302C6 map 10, LOW-PASS SEQUENCE SAMPLING. |          |     |        |                 |

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominiidae; Homo.  
1 (bases 1 to 77050)

REFERENCE

**AUTHORS** Birren, B., Linton, L., Nusbaum, C. and Lander, E.  
**TITLE** Homo sapiens chromosome 10, clone RP11-302C6

Unpublished  
2 (bases 1 to 77060)

Birren, B., Linton, L., Nusbaum, C., Lander, E., Abraham, H., Allen, N., Anderson, S., Baldwin, J., Barna, N., Bastien, V., Beda, F., Boguslavsky, L., Boukhalter, B., Brown, A., Burkett, G., Campionaro, A., Castle, A., Choepel, Y., Colangelo, M., Collins, S., Collymore, A., Cooke, P., DeAtrillano, K., Dewar, K., Diaz, J. S., Dodge, S., Domino, M., Doyle, M., Ferreira, P., FitzHugh, W., Gage, D., Galagan, J., Gardyna, S., Ginde, S., Goyette, M., Graham, L., Grand-Pierre, N., Grant, G., Hagos, B., Heaford, A., Horton, L., Howland, J. C., Iliev, I., Johnson, R., Jones, C., Kann, L., Karatas, A., Klein, J., Laroque, K., Lamarques, R., Landers, T., Lehocsky, J., Levine, R., Lieu, C., Liu, G., Locke, K., Macdonald, P., Marquis, N., McCarthy, M., McGowan, P., McGurk, A., McKernan, K., McPheeters, R., Meldrum, J., Meneus, L., Mihova, T., Miranda, C., Mieng, V., Morrow, J., Murphy, T., Naylor, J., Norman, C. H., O'Connor, T., O'Donnell, P., O'Neil, D., Oliver, T. M., Oliver, J., Peterson, K., Pierre, N., Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D., Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B., Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J., Tesfaye, S., Theodore, J., Tirrell, A., Travers, M., Trigglio, J., Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W. J., Young, G., Zainoun, J., Zimmer, A. and Zody, M.

|         |  |
|---------|--|
| TITLE   | Direct Submission  |
| JOURNAL | Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA   |
| COMMENT | All repeats were identified using RepeatMasker:<br>Smit, A.F.A. & Green, P. (1996-1997)<br><a href="http://ftp.genome.washington.edu/RM/RepeatMasker.html">http://ftp.genome.washington.edu/RM/RepeatMasker.html</a> |

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: <http://www-seq.wi.mit.edu>  
Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)  
----- project Information  
Center project name: L8116  
Center clone name: 302 C 6

\* NOTE: This record contains 86 individual  
\* sequencing reads that have not been assembled into  
\* contigs. Runs of N are used to separate the reads  
\* and the order in which they appear is completely  
\* arbitrary. Low-pass sequence sampling is useful for  
\* identifying clones that may be gene-rich and allows

\* 30554 31339: contig of 786 bp in length  
 \* 31340 31439: gap of 100 bp  
 \* 31440 32242: contig of 803 bp in length  
 \* 32243 32342: gap of 100 bp  
 \* 32343 33137: contig of 795 bp in length  
 \* 33138 33237: gap of 100 bp  
 \* 33238 34042: contig of 805 bp in length  
 \* 34043 34142: gap of 100 bp  
 \* 34143 34961: contig of 819 bp in length  
 \* 34962 35061: gap of 100 bp  
 \* 35062 35863: contig of 802 bp in length  
 \* 35864 35963: gap of 100 bp  
 \* 35964 36767: contig of 804 bp in length  
 \* 36768 36867: gap of 100 bp  
 \* 36868 37646: contig of 779 bp in length  
 \* 37647 37746: gap of 100 bp  
 \* 37747 38559: contig of 813 bp in length  
 \* 38560 38659: gap of 100 bp  
 \* 38660 39458: contig of 799 bp in length  
 \* 39459 39558: gap of 100 bp  
 \* 39559 40371: contig of 813 bp in length  
 \* 40372 40471: gap of 100 bp  
 \* 40472 41210: contig of 739 bp in length  
 \* 41211 41310: gap of 100 bp  
 \* 41311 42103: contig of 793 bp in length  
 \* 42104 42203: gap of 100 bp  
 \* 42204 43003: contig of 800 bp in length  
 \* 43004 43103: gap of 100 bp  
 \* 43104 43905: contig of 802 bp in length  
 \* 43906 44005: gap of 100 bp  
 \* 44006 44825: contig of 820 bp in length  
 \* 44826 44925: gap of 100 bp  
 \* 44926 45722: contig of 797 bp in length  
 \* 45723 45822: gap of 100 bp  
 \* 45823 46629: contig of 807 bp in length  
 \* 46630 46729: gap of 100 bp  
 \* 46730 47549: contig of 820 bp in length  
 \* 47550 47649: gap of 100 bp  
 \* 47650 48440: contig of 791 bp in length  
 \* 48441 48540: gap of 100 bp  
 \* 48541 49343: contig of 803 bp in length  
 \* 49344 49443: gap of 100 bp  
 \* 49444 50256: contig of 813 bp in length  
 \* 50257 50356: gap of 100 bp  
 \* 50357 51159: contig of 803 bp in length  
 \* 51160 51259: gap of 100 bp  
 \* 51260 52054: contig of 795 bp in length  
 \* 52055 52154: gap of 100 bp  
 \* 52155 52934: contig of 780 bp in length  
 \* 52935 53034: gap of 100 bp  
 \* 53035 53830: contig of 796 bp in length  
 \* 53831 53930: gap of 100 bp  
 \* 53931 54736: contig of 806 bp in length  
 \* 54737 54836: gap of 100 bp  
 \* 54837 55639: contig of 803 bp in length  
 \* 55640 55739: gap of 100 bp  
 \* 55740 56549: contig of 810 bp in length  
 \* 56550 56649: gap of 100 bp  
 \* 56650 57458: contig of 809 bp in length  
 \* 57459 57558: gap of 100 bp  
 \* 57559 58341: contig of 783 bp in length  
 \* 58342 58441: gap of 100 bp  
 \* 58442 59258: contig of 817 bp in length  
 \* 59259 59358: gap of 100 bp  
 \* 59359 60095: contig of 737 bp in length  
 \* 60096 60195: gap of 100 bp  
 \* 60196 60999: contig of 804 bp in length  
 \* 61000 61099: gap of 100 bp  
 \* 61100 61913: contig of 814 bp in length  
 \* 61914 62013: gap of 100 bp

Query Match 37.1%; Score 371.6; DB 2; Length 77060;  
 Best Local Similarity 83.5%; Pred. No. 3.1e-80;

Matches 471; Conservative 0; Mismatches 84; Indels 9; Gaps 4;  
 QY 8 TCAGTTAGGAACAATAAGTGCAGAGATCTATTGTACTTTGGTGCATACAGTTAAATCTATT 67  
 DB 36571 TTAGACAAGAGGAATAAATAACAACAGGTTTATTGGATTTGGTGACTACATTTAATGTACT 36512  
 QY 68 GTGTTCTTTGAC-----TAATACAGTAGATTTCGAGTGTTCACAAACAACAAACATGATG 121  
 DB 36511 CTGTTCTTGAACATCATATAAGACAGTAGATTTTGAGTGTTCACACAGCAAAAATGATA 36452  
 QY 122 GGTATGTGAGGTAAATGCATATGCAAACTAGCTTGGGTTAAACCATCCCAATATGTGTGT 181  
 DB 36451 GGTATGTGAGGTAAATGCATATGCCAATAGCTTGGGTTAAACCATCCCAAGTGTGTGCAT 36392  
 QY 182 ATTTCAAAAACAGTACCATAAAATGCAGACAATTTTGTGTCAGTTACAATCAAAAAGTTT 241  
 DB 36391 GTTTCAAAAACGGTACCATAAAATGTAGACAATTTT-TATCAGTTACAAT-AAAAAGTTT 36334  
 QY 242 AAAATGAGGACCTTAGGGTGGTCTCTTAATCCCAATCTAAGTGTATGTCTCCATGAAAGAGA 301  
 DB 36333 AAAATGAAGGCCTTAGGGTGGGCCCAAAATCCAATGTAACTGTCTCCATGAAAGAGA 36274  
 QY 302 AATAAGGATACAAATGTGCACACAGAGAAATGGCCACATGAGGACACAAATGAGATGT 361  
 DB 36273 TATAAGGATACAAATGTGTACATACAGAGAAATGGCCACATGAAAGAGAGAAGT 36214  
 QY 362 GGCTACTTACAAGCCTTAGGAGAGAGGCCCTCCGAGAAAACACACACCTTACCACACCTT 421  
 DB 36213 GCCCACTTACAAGCCAGGAGAGATGTCT-CGGGGAAACCCCTCCCTGCTCATGTTTGTAC 36155  
 QY 422 GTTGCACTTCACTCTGTAGACGAAGTCCCTCCACCTCTTCATCATCGTGGAGGCCCTTTCGA 481  
 DB 36154 CTTTGACATCATCTCCAGATAAAGTCTTCATCTCTCTTCATCATAGATGAGTCTCTCA 36095  
 QY 482 TTCTGAATATCTCCAAATGCTGGAGGTACAAAAGTGAAGAGACAGACAGACACCTCAGG 541  
 DB 36094 TCCTGAATACTCTCCAAATTTTGGAGGTACGAACTGAGAGGACAGACAGACACCTCAGG 36035  
 QY 542 GTGAAAAGTTTAAAGAGATAACA 565  
 DB 36034 GTGAAAAGTTTAAAGAGATAACA 36011  
 RESULT 9  
 LOCUS AC025998 77060 bp DNA linear HTG 19-MAR-2000  
 DEFINITION Homo sapiens chromosome 10 clone RP11-302C6 map 10, LOW-PASS  
 SEQUENCE SAMPLING.  
 ACCESSION AC025998  
 VERSION AC025998.1 GI:7263216  
 KEYWORDS HTG; HTGS\_PHASE0.  
 SOURCE Homo sapiens.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
 Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 REFERENCE 1 (bases 1 to 77060)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C. and Lander,E.  
 TITLE Homo sapiens chromosome 10, clone RP11-302C6  
 JOURNAL Unpublished  
 REFERENCE 2 (bases 1 to 77060)  
 AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,  
 Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,  
 Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,  
 Campoliano,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,  
 Collumore,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S.,  
 Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,  
 Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,  
 Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,  
 Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,  
 Klein,J., Lacombe,K., Lamazares,R., Landers,T., Lehoczyk,J.,  
 Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,  
 McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,  
 Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,

Murphy, T., Naylor, J., Norman, C.H., O'Connor, T., O'Donnell, P.,  
 O'Neil, D., Oliver, T.M., Oliver, J., Peterson, K., Pierre, N.,  
 Pisani, C., Pollara, V., Raymond, C., Riley, R., Rogov, P., Rothman, D.,  
 Roy, A., Santos, R., Schauer, S., Severy, P., Spencer, B.,  
 Stange-Thomann, N., Stojanovic, N., Subramanian, A., Talamas, J.,  
 Testaye, S., Theodore, J., Tirrell, A., Travers, M., Triggillo, J.,  
 Vassiliev, H., Viel, R., Vo, A., Wilson, B., Wu, X., Wyman, D., Ye, W.J.,  
 Young, G., Zainoun, J., Zimmer, A. and Zody, M.

# TITLE JOURNAL

## COMMENT

Submitted (19-MAR-2000) Whitehead Institute/MIT Center for Genome

Research, 320 Charles Street, Cambridge, MA 02141, USA

All repeats were identified using RepeatMasker:

Smit, A.F.A. & Green, P. (1996-1997)

<http://ftp.genome.washington.edu/RM/RepeatMasker.html>

----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research

Center code: WIBR

Web site: <http://www-seq.wi.mit.edu>

Contact: [sequence\\_submissions@genome.wi.mit.edu](mailto:sequence_submissions@genome.wi.mit.edu)

----- Project Information

Center project name: L8116

Center clone name: 302\_C\_6

\* NOTE: This record contains 86 individual  
 \* sequencing reads that have not been assembled into  
 \* contigs. Runs of N are used to separate the reads  
 \* and the order in which they appear is completely  
 \* arbitrary. Low-pass sequence sampling is useful for  
 \* identifying clones that may be gene-rich and allows  
 \* overlap relationships among clones to be deduced.  
 \* However, it should not be assumed that this clone  
 \* will be sequenced to completion. In the event that  
 \* the record is updated, the accession number will  
 \* be preserved.

\* 1 773: contig of 773 bp in length  
 \* 774 873: gap of 100 bp  
 \* 874 1684: contig of 811 bp in length  
 \* 1685 1784: gap of 100 bp  
 \* 1785 2572: contig of 788 bp in length  
 \* 2573 2672: gap of 100 bp  
 \* 2673 3467: contig of 795 bp in length  
 \* 3468 3567: gap of 100 bp  
 \* 3568 4354: contig of 787 bp in length  
 \* 4355 4454: gap of 100 bp  
 \* 4455 5254: contig of 800 bp in length  
 \* 5255 5354: gap of 100 bp  
 \* 5355 6140: contig of 786 bp in length  
 \* 6141 6240: gap of 100 bp  
 \* 6241 7045: contig of 805 bp in length  
 \* 7046 7145: gap of 100 bp  
 \* 7146 7932: contig of 787 bp in length  
 \* 7933 8032: gap of 100 bp  
 \* 8033 8851: contig of 819 bp in length  
 \* 8852 8951: gap of 100 bp  
 \* 8952 9754: contig of 803 bp in length  
 \* 9755 9854: gap of 100 bp  
 \* 9855 10640: contig of 786 bp in length  
 \* 10641 10740: gap of 100 bp  
 \* 10741 11547: contig of 807 bp in length  
 \* 11548 11647: gap of 100 bp  
 \* 11648 12448: contig of 801 bp in length  
 \* 12449 12548: gap of 100 bp  
 \* 12549 13346: contig of 798 bp in length  
 \* 13347 13446: gap of 100 bp  
 \* 13447 14234: contig of 788 bp in length  
 \* 14235 14334: gap of 100 bp  
 \* 14335 15106: contig of 772 bp in length  
 \* 15107 15206: gap of 100 bp  
 \* 15207 16013: contig of 807 bp in length  
 \* 16014 16113: gap of 100 bp  
 \* 16114 16924: contig of 811 bp in length  
 \* 16925 17024: gap of 100 bp  
 \* 17025 17798: contig of 774 bp in length

\* 17799 17898: gap of 100 bp  
 \* 17899 18706: contig of 808 bp in length  
 \* 18707 18806: gap of 100 bp  
 \* 18807 19616: contig of 810 bp in length  
 \* 19617 19716: gap of 100 bp  
 \* 19717 20523: contig of 807 bp in length  
 \* 20524 20623: gap of 100 bp  
 \* 20624 21400: contig of 777 bp in length  
 \* 21401 21500: gap of 100 bp  
 \* 21501 22285: contig of 785 bp in length  
 \* 22286 22385: gap of 100 bp  
 \* 22386 23185: contig of 800 bp in length  
 \* 23186 23286: gap of 100 bp  
 \* 23286 24103: contig of 818 bp in length  
 \* 24104 24203: gap of 100 bp  
 \* 24204 25012: contig of 809 bp in length  
 \* 25013 25112: gap of 100 bp  
 \* 25113 25922: contig of 810 bp in length  
 \* 25923 26022: gap of 100 bp  
 \* 26023 26839: contig of 817 bp in length  
 \* 26840 26939: gap of 100 bp  
 \* 26940 27765: contig of 826 bp in length  
 \* 27766 27865: gap of 100 bp  
 \* 27866 28662: contig of 797 bp in length  
 \* 28663 28762: gap of 100 bp  
 \* 28763 29574: contig of 812 bp in length  
 \* 29575 29674: gap of 100 bp  
 \* 29675 30453: contig of 779 bp in length  
 \* 30454 30553: gap of 100 bp  
 \* 30554 31339: contig of 786 bp in length  
 \* 31340 31439: gap of 100 bp  
 \* 31440 32242: contig of 803 bp in length  
 \* 32243 32342: gap of 100 bp  
 \* 32343 33137: contig of 795 bp in length  
 \* 33138 33237: gap of 100 bp  
 \* 33238 34042: contig of 805 bp in length  
 \* 34043 34142: gap of 100 bp  
 \* 34143 34961: contig of 819 bp in length  
 \* 34962 35061: gap of 100 bp  
 \* 35062 35863: contig of 802 bp in length  
 \* 35864 35963: gap of 100 bp  
 \* 35964 36767: contig of 804 bp in length  
 \* 36768 36867: gap of 100 bp  
 \* 36868 37646: contig of 779 bp in length  
 \* 37647 37746: gap of 100 bp  
 \* 37747 38559: contig of 813 bp in length  
 \* 38560 38659: gap of 100 bp  
 \* 38660 39458: contig of 799 bp in length  
 \* 39459 39558: gap of 100 bp  
 \* 39559 40371: contig of 813 bp in length  
 \* 40372 40471: gap of 100 bp  
 \* 40472 41210: contig of 739 bp in length  
 \* 41211 41310: gap of 100 bp  
 \* 41311 42103: contig of 793 bp in length  
 \* 42104 42203: gap of 100 bp  
 \* 42204 43003: contig of 800 bp in length  
 \* 43004 43103: gap of 100 bp  
 \* 43104 43905: contig of 802 bp in length  
 \* 43906 44005: gap of 100 bp  
 \* 44006 44825: contig of 820 bp in length  
 \* 44826 44925: gap of 100 bp  
 \* 44926 45722: contig of 797 bp in length  
 \* 45723 45822: gap of 100 bp  
 \* 45823 46629: contig of 807 bp in length  
 \* 46630 46729: gap of 100 bp  
 \* 46730 47549: contig of 820 bp in length  
 \* 47550 47649: gap of 100 bp  
 \* 47650 48440: contig of 791 bp in length  
 \* 48441 48540: gap of 100 bp  
 \* 48541 49343: contig of 803 bp in length  
 \* 49344 49443: gap of 100 bp  
 \* 49444 50256: contig of 813 bp in length  
 \* 50257 50356: gap of 100 bp

|   |  |   |
|---|--|---|
| Query Match   |  | 36 58; Score 365.2; DB 2; Length 77060; |
| Best Local Similarity   |  | 69.08; Pred. No. 1.le 769;              |
| Matches 542; Conservative 0; Mismatches 209; Indels 35; Gaps 8; |  |   |
| Qy  | 8 TCAGTTAGGAAGTAAGTCAAGAGATCTATTGTACTTGGTGTGACTACAGTTAATGTATT        | 67                                      |
| Db  | 22530 TTAGCAGAGGAATAAGTCAAGAGATCTATTGTACTTGGTGTGACTACAGTTAATGTATT    | 22589                                   |
| Qy  | 68 GTGTTCTTGA-----CTAATACAGTAGATTTCGAGTGTCTTCACACAAAACATGATG         | 121                                     |
| Db  | 22590 CTGTTCTTGAACATGCTAGGACAGATAGATTTCGAGTGTCTTCACACAAAACATGATG     | 22648                                   |
| Qy  | 122 GGTATGTGAGTAAATGTCACAACTAGCTTGGTTAAACATTCACCAATATGTGTGT          | 181                                     |
| Db  | 22649 GGTGTGAGGTAATGTCATATGCGCATTTCTTGGTTAAACATTCACCAATATGTGTGT      | 22708                                   |
| Qy  | 182 ATTTCAAAACAGTACCATAAATGACAGACAATTTGTGTGCTAGTTACAAATCAAAAGTTT     | 241                                     |
| Db  | 22709 ATTTCAACACAGTACCATAAATGAGGATGTTT-TATCAGTTACA-----AAAAAAATTT    | 22764                                   |
| Qy  | 242 AAATGAGGACCTTAGGTGGTCTTAATCCAAATTAAGTATGATGTCTCCATGAAGAGGA       | 301                                     |
| Db  | 22765 TAAATGAGGACCTTAGGAGAGAACCTTAATCCAATGTAATCTGTCTCCACAAAGGAGGA    | 22824                                   |
| Qy  | 302 AATAAGGATACAAATGTCACACAGAGAGAAATGCCACATGAGGACACCAATGAGATGT       | 361                                     |
| Db  | 22825 GATTAAGATACAAAGTCCACACAGAGAAATGCCACATGAGGACATAGAGGAGAGGC       | 22884                                   |
| Qy  | 362 GGCTACTTTACAGCCTAGGAGAGGCGCTCCGAGAAAAACACACCCCTACCCACCTTGAT      | 421                                     |
| Db  | 22885 GGCCACTTTACAAACCCAGGAGAGGCGCTCAGAGAAAAACCACTTGGCCCATCTTGAT     | 22944                                   |
| Qy  | 422 GTTGGACTTTCATCTGTAGAGCAAGTCTCCACCCCTCTTCATCAGGTGGAAGCCTTTGA      | 481                                     |
| Db  | 22945 CTTGGACTTTCAT-----CTCCCTTCATCAAAATGGAGGCTTCTGA                 | 22983                                   |
| Qy  | 482 TCTTGAATATTTCTCAAAATGCTGGAAGGTACAAAAGTCAAGAGACAGACAGACCTCAGG     | 541                                     |
| Db  | 22984 TCCTGAATACTTTCCAAAACCTAGAAAGTACAGTGGTGAAGGAGGAGGAGGAGCTCANG    | 23043                                   |
| Qy  | 542 GTGAAAAGTTTAAAGAGAAATACATCTTTCCATTGCTGTGCTCTATCCCTTACACACACC     | 601                                     |
| Db  | 23044 GTGAAAAGTTTAAAGAG-ATAACACCTCCCATTTGCCCTGTGCTCTATCCCTAACAC-CACC | 23101                                   |
| Qy  | 602 TATTCCAGTCTTTATTTGTTGTTTTCGTTGTTCTGCTGGGTATAGTGTAGTTGTGTA        | 661                                     |

|  |  |                                      |
|--|--|--------------------------------------|
| Db   | 23102 TGTGCCAGCCTTTATTGTGTTTTCATCTCCCTGTCTGGGATAGTGTAACTTNNNT  | 23161                                |
| Qy   | 662 TCTGTGT--TTACATACAGGATAAACAACAAGGTAACAATAAATAAACAAGACA   | 720                                  |
| Db   | 23162 GATTNGTGTTCATAGATGATAACNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNNN  | 23221                                |
| Qy   | 721 GCAAAACTCAACTAATAGTGTGTTGGGCATGTTGACAGTGAAGACAGAGAGTCACATAA  | 780                                  |
| Db   | 23222 NNN  | 23281                                |
| Qy   | 781 AATGGA 786   |                                      |
| Db   | 23282 NNNNGA 23287   |                                      |
| RESULT 10  |  |                                      |
| AC005725/c   |  |                                      |
| LOCUS  | AC005725   | 197915 bp DNA linear PRI 02-DEC-1998 |
| DEFINITION   | Homo sapiens chromosome 17, clone hRPK.680_C_21, complete sequence.  |                                      |
| ACCESSION  | AC005725   |                                      |
| VERSION  | AC005725.1   | GI:3947428                           |
| KEYWORDS   | HTG.   |                                      |
| SOURCE   | Homo sapiens.  |                                      |
| ORGANISM   | Homo sapiens   |                                      |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo. |  |                                      |
| REFERENCE  | 1 (bases 1 to 197915)  |                                      |
| AUTHORS  | Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collumore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K., Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S., Geraghty,K., Grant,G., Hagos,B., Heaford,A., Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L., Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Nahr,R., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A., Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C., Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D., Ye,W.J., Zhao,J. and Zody,M. |                                      |
| TITLE  | Direct Submission  |                                      |
| JOURNAL  | Submitted (29-SEP-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA   |                                      |
| REFERENCE  | 3 (bases 1 to 197915)  |                                      |
| AUTHORS  | Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boutwell,C., Brown,A., Castle,A., Cerny,J., Colangelo,M., Collins,S., Collumore,A., Cooke,P., Corliss,D., Depayre,E., Devon,K., Dewar,K., Donelan,L., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S., Geraghty,K., Grant,G., Hagos,B., Heaford,A., Herena,L., Horton,L., Howland,J.C., Jacotot,L., Jones,C., Kann,L., Karatas,A., Lehoczy,J., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J., Nahr,R., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A., Severy,P., Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C., Subramanian,A., Tesfaye,S., Tichovolsky,N., Torruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D., Ye,W.J., Zhao,J. and Zody,M. |                                      |
| TITLE  | Direct Submission  |                                      |
| JOURNAL  | Submitted (02-DEC-1998) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA   |                                      |
| COMMENT  | On Dec 2, 1998 this sequence version replaced gi:393522. All repeats were identified using RepeatMasker: Smit, A.F.A. & Green, P. (1996-1997) http://ftp.genome.washington.edu/RM/RepeatMasker.html  |                                      |

Only the last 197915 base pairs of this clone are being submitted.  
The remainder overlaps accession number AC005548 (WICGR project L361).

| FEATURES      |  | Location/Qualifiers                    |
|---------------|--|--|
| source        |  | 1..197915                              |
|               |  | /organism="Homo sapiens"               |
|               |  | /db_xref="taxon:9606"                  |
|               |  | /chromosome="17"                       |
|               |  | /map="17"                              |
|               |  | /clone="hRPK_680_C-21"                 |
|               |  | /clone_lib="RPCI-11 human BAC library" |
| repeat_region |  | complement(251..355)                   |
| repeat_region |  | /rpt_family="MIR"                      |
| repeat_region |  | 687..969                               |
| repeat_region |  | /rpt_family="AluSc"                    |
| repeat_region |  | 2640..2661                             |
| repeat_region |  | /rpt_family="AT_rich"                  |
| repeat_region |  | 4789..5192                             |
| repeat_region |  | /rpt_family="MLT1B"                    |
| repeat_region |  | complement(5330..5795)                 |
| repeat_region |  | /rpt_family="MLT1H"                    |
| repeat_region |  | complement(7650..7676)                 |
| repeat_region |  | /rpt_family="TGAAn"                    |
| repeat_region |  | complement(7722..8030)                 |
| repeat_region |  | /rpt_family="LTR16A"                   |
| repeat_region |  | complement(8051..8475)                 |
| repeat_region |  | /rpt_family="L2"                       |
| repeat_region |  | complement(8580..8676)                 |
| repeat_region |  | /rpt_family="MIR"                      |
| repeat_region |  | complement(8866..8935)                 |
| repeat_region |  | /rpt_family="MLT1I"                    |
| repeat_region |  | 8936..8995                             |
| repeat_region |  | /rpt_family="TAGAn"                    |
| repeat_region |  | complement(8998..9054)                 |
| repeat_region |  | /rpt_family="TAGAn"                    |
| repeat_region |  | complement(9057..9381)                 |
| repeat_region |  | /rpt_family="MLT1I"                    |
| repeat_region |  | 9391..9691                             |
| repeat_region |  | /rpt_family="AluSx"                    |
| repeat_region |  | 9756..9824                             |
| repeat_region |  | /rpt_family="polypurine"               |
| repeat_region |  | complement(9825..9847)                 |
| repeat_region |  | /rpt_family="MLT1J"                    |
| repeat_region |  | 9848..9966                             |
| repeat_region |  | /rpt_family="GAAAn"                    |
| repeat_region |  | complement(9967..10060)                |
| repeat_region |  | /rpt_family="MLT1J"                    |
| repeat_region |  | 10947..11128                           |
| repeat_region |  | /rpt_family="L2"                       |
| repeat_region |  | 11187..11570                           |
| repeat_region |  | /rpt_family="L2"                       |
| repeat_region |  | complement(12968..13015)               |
| repeat_region |  | /rpt_family="(CGG)n"                   |
| repeat_region |  | 13110..13159                           |
| repeat_region |  | /rpt_family="GC_rich"                  |
| repeat_region |  | complement(13396..13420)               |
| repeat_region |  | /rpt_family="(CGG)n"                   |
| repeat_region |  | complement(13586..13654)               |
| repeat_region |  | /rpt_family="(CGG)n"                   |
| repeat_region |  | 13667..13758                           |
| repeat_region |  | /rpt_family="GC_rich"                  |
| repeat_region |  | 13884..13927                           |
| repeat_region |  | /rpt_family="GC_rich"                  |
| repeat_region |  | 13973..14023                           |
| repeat_region |  | /rpt_family="GC_rich"                  |
| repeat_region |  | 18546..18848                           |
| repeat_region |  | /rpt_family="AluY"                     |
| repeat_region |  | complement(20108..20409)               |
| repeat_region |  | /rpt_family="AluY"                     |
| repeat_region |  | complement(20687..20805)               |
| repeat_region |  | /rpt_family="MIR"                      |
| repeat_region |  | 21194..21495                           |
| repeat_region |  | /rpt_family="AluY"                     |
| repeat_region |  | complement(21620..21682)               |
| repeat_region |  | /rpt_family="(CA)n"                    |
| repeat_region |  | complement(23260..23416)               |
| repeat_region |  | /rpt_family="MIR"                      |
| repeat_region |  | 26556..26614                           |
| repeat_region |  | /rpt_family="MIR"                      |
| repeat_region |  | 27378..27513                           |
| repeat_region |  | /rpt_family="L2"                       |
| repeat_region |  | 27644..27727                           |
| repeat_region |  | /rpt_family="purine-rich"              |
| repeat_region |  | complement(28209..28257)               |
| repeat_region |  | /rpt_family="L2"                       |
| repeat_region |  | complement(28225..28485)               |
| repeat_region |  | /rpt_family="MIR"                      |
| repeat_region |  | 28549..28847                           |
| repeat_region |  | /rpt_family="AluSx"                    |
| repeat_region |  | 29715..30026                           |
| repeat_region |  | /rpt_family="AluSg"                    |
| repeat_region |  | 30027..30068                           |
| repeat_region |  | /rpt_family="polypurine"               |
| repeat_region |  | 30328..30353                           |
| repeat_region |  | /rpt_family="AT_rich"                  |
| repeat_region |  | 31266..31388                           |
| repeat_region |  | /rpt_family="(TGA)n"                   |
| repeat_region |  | 31474..31520                           |
| repeat_region |  | /rpt_family="purine-rich"              |
| repeat_region |  | 31537..31784                           |
| repeat_region |  | /rpt_family="MIR"                      |
| repeat_region |  | complement(33067..33188)               |
| repeat_region |  | /rpt_family="purine-rich"              |
| repeat_region |  | complement(33831..34379)               |
| repeat_region |  | /rpt_family="L1MB6"                    |
| repeat_region |  | complement(34377..34527)               |
| repeat_region |  | /rpt_family="L1"                       |
| repeat_region |  | complement(36437..36503)               |
| repeat_region |  | /rpt_family="MIR"                      |
| repeat_region |  | complement(36546..36745)               |
| repeat_region |  | /rpt_family="MIR"                      |
| repeat_region |  | 37422..37595                           |
| repeat_region |  | /rpt_family="MERSB"                    |
| repeat_region |  | complement(37647..37836)               |
| repeat_region |  | /rpt_family="MERSB"                    |
| repeat_region |  | 38648..38746                           |
| repeat_region |  | /rpt_family="MERSA"                    |
| repeat_region |  | 38818..39002                           |
| repeat_region |  | /rpt_family="MIR"                      |
| repeat_region |  | complement(39298..39373)               |
| repeat_region |  | /rpt_family="MIR"                      |
| repeat_region |  | 39598..39749                           |
| repeat_region |  | /rpt_family="MERSA"                    |
| repeat_region |  | complement(39764..39885)               |
| repeat_region |  | /rpt_family="MERSB"                    |
| repeat_region |  | 40689..41149                           |
| repeat_region |  | /rpt_family="L2"                       |
| repeat_region |  | complement(41364..41675)               |
| repeat_region |  | /rpt_family="AluJb"                    |
| repeat_region |  | 42215..42272                           |
| repeat_region |  | /rpt_family="MER3"                     |
| repeat_region |  | complement(42297..42607)               |
| repeat_region |  | /rpt_family="AluSg"                    |
| repeat_region |  | complement(42851..43091)               |
| repeat_region |  | /rpt_family="MERS8A"                   |
| repeat_region |  | 43312..43376                           |
| repeat_region |  | /rpt_family="MLT2D"                    |
| repeat_region |  | complement(43767..44095)               |
| repeat_region |  | /rpt_family="AluSg"                    |
| repeat_region |  | 44800..44932                           |
| repeat_region |  | /rpt_family="MIR"                      |
| repeat_region |  | complement(45393..45589)               |
| repeat_region |  | /rpt_family="MIR"                      |
| repeat_region |  | complement(47813..48306)               |

Query Match

23.4%; Score 234.4; DB 9; Length 197915;

Best Local Similarity 73.3%; Pred. No. 7.5e-47;  
Matches 318; Conservative 0; Mismatches 106; Indels 10; Gaps 1;

QY 8 TCAGTTAGGAAGAAATAGTCGACAGATCTATTGTGCTGACACAGTTAATGCTATT 67  
DB 64448 TTAGCAGGAGAGAAATAGTCGACAGATCTATTGAAATATGTTCT-----GTTTC 64399  
QY 68 GTGTTCTTGACTAATACAGTAGATTTCGAGTGTCTTCACAAACAAAACATGATGGGTATG 127  
DB 64398 TTGAACGCTGTACAGCAGTAGACTTTTCAGTGTCTTCACAAACAAAAGATGATGGGTATG 64339  
QY 128 TGAGTAAATGACATGCAAACTAGCTTGGGTTAACTTCCAAATATGTGTGTTATTTCA 187  
DB 64338 TGAGTAAACGATATGCTCATGCTGGTGGTTAGCCATTCACACAGCATGATTTCA 64279  
QY 188 AAACAGTACCAATAATGACAGCAATTTGTGTCAGTTACAAATCAAAAAAGTTTAAATG 247  
DB 64278 GAACAGTACCAATAATGAGAAATTTTATCAGTTACAAATAAATAGTTTAAAT 64219  
QY 248 AGGACCTTAGGTGGTCTCTATCCAAATCTAAGTGTCTCTCATGAAAGAGAAATAAG 307  
DB 64218 GAGGACCTCGATGGCCCTATCTAGTCTCAACTGATGCTCCCTTAAAGAGAGGTAG 64159  
QY 308 GATACAAATGTGCACACAGAGAGAAATGGCCACATGAGGACACAAATGAGATGTGGCTAC 367  
DB 64158 GATACAGATGTGCACACACAGAGGGGTGACCAATGAGGACACAGGGAGGGCGCCAC 64099  
QY 368 TTACAGCCTAGGAGAGGGCTCCGAGAAACACACACCTACCCACACCTTGATTTGGA 427  
DB 64098 CTACAGTCCGGGAGAGGGCTCAGAAGAACCCACCTGCCACATCTTGATTTGGA 64039  
QY 428 CTTCACTCTGTAGA 441  
DB 64038 TTCAACCTTCAGA 64025

RESULT 11  
AC025281 193259 bp DNA linear PRI 30-MAR-2002  
LOCUS Homo sapiens chromosome 16 clone RP11-356024, complete sequence.  
DEFINITION AC025281  
ACCESSION AC025281.7 GI:19848339  
VERSION HTG.  
KEYWORDS Homo sapiens.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE 1 (bases 1 to 193259)  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE DOE Joint Genome Institute.  
JOURNAL Sequencing of Human Chromosome 16  
REFERENCE 2 (bases 1 to 193259)  
Unpublished  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (08-MAR-2000) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE 3 (bases 1 to 193259)  
DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (30-MAR-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT On Mar 30, 2002 this sequence version replaced gi:13786380.  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.  
-----  
Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.  
-----  
FEATURES  
source Location/Qualifiers  
1. 193259  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-356024"  
BASE COUNT 53957 a 40929 c 42006 g 56367 t  
ORIGIN  
Query Match 10.1%; Score 101.4; DB 9; Length 193259;  
Best Local Similarity 73.7%; Pred. No. 1.9e-14;  
Matches 157; Conservative 0; Mismatches 51; Indels 5; Gaps 2;  
QY 219 TCAGTTACAAATCAAAAAAGTTTAAATAGGACCTTAGGTGGTCTTAATCCAACTA 278  
DB 51145 TCAGTCTCAAAAAAGGTAATAGTCTATTAGGTGGTCTTAATCCAACTA 51204  
QY 279 AGTGATGTCTCATCA-AAGAGGAATAGGATCAAAATGTGCACACAGAGAAATGCG 337  
DB 51205 ACTGATGTGCTAATGAAGAGGAGGACAAAGGACACAGAT----ACACAGAGGAAAGAC 51260  
QY 338 CACATGAGGACACAAATGAGAATGTGCTACTTACAAAGCCTTAGGAGAGGCGCTCCGAGAA 397  
DB 51261 CTTGTGAAGACACAAAGAGAGATGTCATTTACAAGCCAGAGAGACATTCAGAGA 51320  
QY 398 AACACACCTCCACACACCTTGATCTTGACTT 430  
DB 51321 AATCAACCTGTCAACACCTTCATCTTGACTT 51353  
RESULT 12  
AC016716 214269 bp DNA linear PRI 09-MAY-2001  
LOCUS Homo sapiens BAC clone RP11-31213 from 2, complete sequence.  
DEFINITION AC016716  
ACCESSION AC016716  
VERSION AC016716.6 GI:11038573  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens.  
REFERENCE 1 (bases 1 to 214269)  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE Sulston, J.E. and Waterston, R.  
AUTHORS Toward a complete human genome sequence  
JOURNAL Genome Res. 8 (11), 1097-1108 (1998)  
MEDLINE 99063792  
PUBMED 9847074  
REFERENCE 2 (bases 1 to 214269)  
Haakenson, B., Coblitz, B., Hawkins, M., Elliott, G., Abbott, A. and Boyer, E.  
TITLE The sequence of Homo sapiens BAC clone RP11-31213  
JOURNAL Unpublished  
REFERENCE 3 (bases 1 to 214269)  
Waterston, R.H.  
AUTHORS Direct Submission  
TITLE Submitted (04-DEC-1999) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
JOURNAL 4 (bases 1 to 214269)  
Waterston, R.H.  
AUTHORS Direct Submission  
TITLE Submitted (30-OCT-2000) Genome Sequencing Center, Washington University School of Medicine, 4444 Forest Park Parkway, St. Louis, MO 63108, USA  
JOURNAL

```

REFERENCE
AUTHORS
TITLE
JOURNAL
COMMENT
5 (bases 1 to 214269)
Waterston,R.
Direct Submission
Submitted (09-MAY-2001) Department of Genetics, Washington
University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
On Oct 30, 2000 this sequence version replaced gi:9910084.
----- Genome Center
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapinsewatson.wustl.edu
----- Summary Statistics
Center project name: H_NH0312I03
-----

```





[illegible]

BASE COUNT 54946 a 42867 c 42439 g 55210 t 8 others  
ORIGIN  
Query Match 9.6%; Score 96.2; DB 2; Length 195470;  
Best Local Similarity 76.2%; Pred. No. 3.5e-13;  
Matches 147; Conservative 0; Mismatches 38; Indels 8; Gaps 2;  
QY 8 TCAGTTAGGAGAATAAGTGCAGAGATCTATTGTACTTGGTGTGACTACAGTTAATGTATT 67  
Db 110474 TTAGACAGAGGAATAAGTTCAAGAGATATACTGTACAACATGATGATTAATGTATT 110533  
QY 68 GGTCTTCTG-----ACTAATACAGTATGTTTCGAGTGTCTCACAAC--AAAAACATGA 119  
Db 110534 GTATTCTTTAAATCACAAGGGGTAGATTTTAAGTGTCTTCACACCAAAAAAATGG 110593  
QY 120 TGGGTATGTGAGGTAAATGCATGCAAACTAGCTTGGGTAAACCATTCACCAATATGTGT 179  
Db 110594 TAAAGTATGTGAGGTAAATGCATATGTTAATAGTCAATTTAGCCATTCACCAATGTCTAT 110653  
QY 180 GTATTTCAAAACA 192  
Db 110654 GTATTTCAAAACA 110666

RESULT 15  
AC093514 147330 bp DNA linear PRI 04-JAN-2002  
LOCUS Homo sapiens chromosome 16 clone CTD-2517022, complete sequence.  
DEFINITION AC093514  
ACCESSION AC093514  
VERSION AC093514.2 GI:18057084  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 147330)  
DOE Joint Genome Institute.  
Sequencing of Human Chromosome 16  
Unpublished  
2 (bases 1 to 147330)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (31-AUG-2001) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
3 (bases 1 to 147330)  
DOE Joint Genome Institute.  
Direct Submission  
Submitted (04-JAN-2002) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On Jan 4, 2002 this sequence version replaced gi:15383803.  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.  
-----  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.  
-----  
FEATURES Location/Qualifiers  
1..147330  
/organism="Homo sapiens"

/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="CTD-2517022"  
BASE COUNT 41780 a 30102 c 30898 g 44550 t  
ORIGIN  
Query Match 9.5%; Score 95.2; DB 9; Length 147330;  
Best Local Similarity 71.9%; Pred. No. 6.3e-13;  
Matches 138; Conservative 0; Mismatches 53; Indels 1; Gaps 1;  
QY 240 TTAATAATGAGGACCTTAGGTGGTCCCTAATCAATCTAAGTGTATCTCCATGAAAGAG 299  
Db 119670 TTAATAATGAGGCCCTTAGGTGGGCCCTAATCAATCTGCTGTGACTTTATGAGAGAG 119729  
QY 300 GAAATAAGGATACAAATGTGCACACAGAGAGAAATGGCCACATGAGGACACAAATGAGAAT 359  
Db 119730 GAAGAGACACCAGAGATGTGCACTCACAGAGGGGGCCATGTGAGGACACAGCAGAGAG 119789  
QY 360 GTGGCTACTTTACAACCCCTAGGAGAGAGGCCCTCCGAGAGAAACACCCCTACCCACACCTTG 419  
Db 119790 GTGGCCATCCGCAAGCAAGGAGAGAGG-CTTCAAGAGAGAACTCTAAACCTGTGTCACCTTG 119848  
QY 420 ATGTTGGAGCTTC 431  
Db 119849 ATCTGGGACTTC 119860

Search completed: December 24, 2002, 21:06:13  
Job time : 3113.19 secs

**THIS PAGE BLANK (uspto)**

GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 16:24:42 ; Search time 1812.02 Seconds  
(without alignments)  
16077.036 Million cell updates/sec

Title: US-09-708-724A-3\_COPY\_70000\_71000  
Perfect score: 1001  
Sequence: 1 ggagatgataaacccgtgtg.....ccaattcaggagtctatgtg 1001

Scoring table: IDENTITY\_NUC  
Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000

Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_om.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID       | Description        |
|------------|-------|-------------|--------|----|----------|--------------------|
| 1          | 1001  | 100.0       | 191768 | 9  | AC092357 | AC092357 Homo sapi |
| 2          | 992.4 | 99.8        | 151049 | 9  | AC018558 | AC018558 Homo sapi |
| c 3        | 836.8 | 83.6        | 191014 | 9  | AC113268 | AC113268 Papio cyn |
| 4          | 835   | 83.4        | 185664 | 9  | AC093249 | AC093249 Homo sapi |
| 5          | 726.8 | 72.6        | 153875 | 9  | AC003682 | AC003682 Homo sapi |
| 6          | 649.2 | 64.9        | 45383  | 9  | AC010642 | AC010642 Homo sapi |
| 7          | 649.2 | 64.9        | 160852 | 9  | AC020915 | AC020915 Homo sapi |
| 8          | 649.2 | 64.9        | 214042 | 2  | AC023149 | AC023149 Homo sapi |
| 9          | 627.4 | 62.7        | 158802 | 2  | AC015589 | AC015589 Homo sapi |
| 10         | 625.8 | 62.5        | 196686 | 9  | AC010328 | AC010328 Homo sapi |
| c 11       | 615   | 61.4        | 124102 | 9  | AC021148 | AC021148 Homo sapi |
| 12         | 615   | 61.4        | 158302 | 2  | AC069481 | AC069481 Homo sapi |
| c 13       | 602   | 60.1        | 4473   | 9  | AF017337 | AF017337 Homo sapi |
| c 14       | 590.2 | 59.0        | 153937 | 9  | AL451075 | AL451075 Human DNA |
| c 15       | 589   | 58.8        | 133704 | 9  | AC008567 | AC008567 Homo sapi |
| c 16       | 584.6 | 58.4        | 131826 | 9  | AC011460 | AC011460 Homo sapi |
| c 17       | 583.6 | 58.3        | 106648 | 2  | AC004588 | AC004588 Homo sapi |
| 18         | 583.6 | 58.3        | 159849 | 2  | AC021454 | AC021454 Homo sapi |
| c 19       | 583.6 | 58.3        | 163915 | 2  | AC087451 | AC087451 Homo sapi |
| c 20       | 583.6 | 58.3        | 166867 | 9  | AP003733 | AP003733 Homo sapi |
| c 21       | 579.6 | 57.9        | 184169 | 2  | AC026349 | AC026349 Homo sapi |
| c 22       | 579.6 | 57.9        | 219553 | 2  | HS312687 | HS312687 Homo sapi |
| c 23       | 574.8 | 57.4        | 160169 | 2  | AC051664 | AC051664 Homo sapi |
| c 24       | 571.6 | 57.1        | 137830 | 9  | AC003029 | AC003029 Homo sapi |
| c 25       | 562   | 56.1        | 196431 | 2  | CNS01DU5 | AL133162 Homo sapi |
| c 26       | 557.2 | 55.7        | 115732 | 9  | AC104566 | AC104566 Homo sapi |
| c 27       | 557.2 | 55.7        | 158535 | 9  | AC093368 | AC093368 Homo sapi |
| c 28       | 557   | 55.6        | 124047 | 9  | AL713966 | AL713966 Human DNA |
| c 29       | 556.4 | 55.6        | 179581 | 2  | AC009397 | AC009397 Homo sapi |
| c 30       | 555.2 | 55.5        | 80611  | 9  | HS323P24 | AL022157 Human DNA |
| c 31       | 553.2 | 55.3        | 35715  | 9  | AC020951 | AC020951 Homo sapi |
| c 32       | 546.4 | 54.6        | 170470 | 9  | AC092364 | AC092364 Homo sapi |
| c 33       | 546.4 | 54.6        | 182154 | 2  | AC041049 | AC041049 Homo sapi |
| c 34       | 544.8 | 54.4        | 129794 | 9  | AL353812 | AL353812 Human DNA |
| c 35       | 544.8 | 54.4        | 132229 | 9  | HS447N6  | 282211 Human DNA S |
| 36         | 544.8 | 54.4        | 171970 | 2  | AL590454 | AL590454 Homo sapi |
| 37         | 544   | 54.3        | 179272 | 2  | AC069149 | AC069149 Homo sapi |
| 38         | 538.8 | 53.8        | 104082 | 9  | AC011503 | AC011503 Homo sapi |
| 39         | 526.8 | 52.6        | 38757  | 9  | AC008976 | AC008976 Homo sapi |
| c 40       | 524.4 | 52.4        | 149794 | 2  | AC007273 | AC007273 Homo sapi |
| c 41       | 524.4 | 52.4        | 161879 | 9  | AC017005 | AC017005 Homo sapi |
| c 42       | 522.8 | 52.2        | 120951 | 9  | AC073649 | AC073649 Homo sapi |
| c 43       | 520.6 | 52.0        | 164295 | 2  | AC025958 | AC025958 Homo sapi |
| 44         | 520.6 | 52.0        | 196826 | 9  | AC026775 | AC026775 Homo sapi |
| c 45       | 520.4 | 52.0        | 187280 | 9  | AC092674 | AC092674 Homo sapi |

ALIGNMENTS

RESULT 1  
AC092357

LOCUS

DEFINITION

AC092357 AC023764

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AC092357 191768 bp DNA linear PRI 15-AUG-2001  
Homo sapiens chromosome 16 clone RP11-332P24, complete sequence.

AC092357.2 GI:15187270

HTG.

Homo sapiens.

Homo sapiens

Eukaryota: Metazoa; Chordata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 191768)

DOE Joint Genome Institute.

Sequencing of Human Chromosome 16

Unpublished

```
REFERENCE
AUTHORS      2 (bases 1 to 191768)
TITLE        DOE Joint Genome Institute.
JOURNAL      Direct Submission
Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE
AUTHORS      3 (bases 1 to 191768)
TITLE        DOE Joint Genome Institute.
JOURNAL      Direct Submission
Submitted (15-AUG-2001) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
On or before Aug 15, 2001 this sequence version replaced
gi:7596817, gi:14589546.

COMMENT
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality
estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than
1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the
GenBank flat file format but are available as part
of this entry's ASN.1 file.
-----

FEATURES
            Location/Qualifiers
source      1..191768
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="16"
            /clone="RP11-332P24"

BASE COUNT  57149 a 39592 c 40150 g 54877 t

Query Match      100.0%; Score 1001; DB 9; Length 191768;
Best Local Similarity 100.0%; Pred. No. 3.6e-300;
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY  1  GGAGATGGATAACCGTGTGAGTGCCTCAAGTGTGTGCGACCATGGAATGGAGACTG 60
|||||
Db  87642 GGAGATGGATAACCGTGTGAGTGCCTCAAGTGTGTGCGACCATGGAATGGAGACTG 87701
|||||

QY  61  GAGGATACATGGATCCCAACTACAGGCCCGCCAGCTCCTCCAGTATGAGCCATGAGCCAGTT 120
|||||
Db  87702 GAGGATACATGGAICCAACTACAGGCCCGCCAGCTCCTCCAGTATGAGCCATGAGCCAGTT 87761
|||||

QY  121 GAATCTGAATGTGAGATGGATGAAGACCGGACGAGTCACTGACGTCAACCCCTCAT 180
|||||
Db  87762 GAATCTGAATGTGAGATGGATGAAGACCGGACGAGTCACTGACGTCAACCCCTCAT 87821
|||||

QY  181 ACATGGGGTCAGATCAAGAAAACACACAGAAAGTGAAGAACTGGTGTAGTCCAGGG 240
|||||
Db  87822 ACATGGGGTCAGATCAAGAAAACACACAGAAAGTGAAGAACTGGTGTAGTCCAGGG 87881
|||||

QY  241 TCAGGCAAAAACCCCTGACTCCTCATGTTTATGGCCATGCTTGTGTAATCTCTGTGCAGT 300
|||||
Db  87882 TCAGGCAAAAACCCCTGACTCCTCATGTTTATGGCCATGCTTGTGTAATCTCTGTGCAGT 87941
|||||

QY  301 ATGATTTTCTGTGCAGAGCAAAAACATATTGGGCAATATTTCTTAACCCACCGGTACT 360
|||||
Db  87942 ATGATTTTCTGTGCAGAGCAAAAACATATTGGGCAATATTTCTTAACCCACCGGTACT 88001
|||||

QY  361 GTGATCATACTCTGAGCAGCACTCCTCCTGAGATATATCATGATCAAGGACATCAGTA 420
|||||
Db  88002 GTGATCATACTCTGAGCAGCACTCCTCCTGAGATATATCATGATCAAGGACATCAGTA 88061
|||||
```

```
QY  421 CCAGACCTTAACCTCCCTCGACACAGAGCAATTAGACTCTCTATAACAATGGTATCAAT 480
|||||
Db  88062 CCAGACCTTAACCTCCCTCGACACAGAGCAATTAGACTCTCTATAACAATGGTATCAAT 88121
|||||

QY  481 TATACCACTCCAGTGGAGGACTTCTTTATGTGTGTCACCCAGGATACATTGCTCAACTGC 540
|||||
Db  88122 TATACCACTCCAGTGGAGGACTTCTTTATGTGTGTCACCCAGGATACATTGCTCAACTGC 88181
|||||

QY  541 AGTTGCCCTTGAGTTGATCCCAAGCATGGTTGAGTTACCAATAAAAAAATATGTACCTA 600
|||||
Db  88182 AGTTGCCCTTGAGTTGATCCCAAGCATGGTTGAGTTACCAATAAAAAAATATGTACCTA 88241
|||||

QY  601 TTAGACCTTAGCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 660
|||||
Db  88242 TTAGACCTTAGCTTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 88301
|||||

QY  661 CCAAAATTTGACTGATATACAGAAATGGGCTCCCTTTGATAATTCACCCCTCCTTGG 720
|||||
Db  88302 CCAAAATTTGACTGATATACAGAAATGGGCTCCCTTTGATAATTCACCCCTCCTTGG 88361
|||||

QY  721 GCCCACTGTCTGGGCCCTTAGCTAGACAATAAGTCCATGTTAATGGGAGACATTATTGAC 780
|||||
Db  88362 GCCCACTGTCTGGGCCCTTAGCTAGACAATAAGTCCATGTTAATGGGAGACATTATTGAC 88421
|||||

QY  781 TGGGTCCTCTGTGTCATTAAAGATGGGAGAGATGAGAATCAGACACATGGCATAACTT 840
|||||
Db  88422 TGGGTCCTCTGTGTCATTAAAGATGGGAGAGATGAGAATCAGACACATGGCATAACTT 88481
|||||

QY  841 CACTGGCACTGGTGGCGAACTTTAAACATCTCTTCACTTCAACACACTGGGATTCATCC 900
|||||
Db  88482 CACTGGCACTGGTGGCGAACTTTAAACATCTCTTCACTTCAACACACTGGGATTCATCC 88541
|||||

QY  901 CAATCTGCATCAACTTGTCTGGCATGGAAGCGGCTTTAGCCACCTTTGGCTCAATGG 960
|||||
Db  88542 CAATCTGCATCAACTTGTCTGGCATGGAAGCGGCTTTAGCCACCTTTGGCTCAATGG 88601
|||||

QY  961 CATTATCAAGGAAGAGAGAGTCCAAATTCAGGAGTCTATGTG 1001
|||||
Db  88602 CATTATCAAGGAAGAGAGAGTCCAAATTCAGGAGTCTATGTG 88642
|||||

RESULT 2
AC018558      Homo sapiens chromosome 16 clone RP11-80F22, complete sequence.
LOCUS         AC018558      151049 bp      DNA      linear      PRI 02-NOV-2001
DEFINITION    Homo sapiens chromosome 16 clone RP11-80F22, complete sequence.
ACCESSION     AC018558
VERSION       AC018558.5  GI:16596530
KEYWORDS      HTG.
SOURCE        Homo sapiens.
ORGANISM      Homo sapiens
              Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
              Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE     1 (bases 1 to 151049)
              DOE Joint Genome Institute.
              Sequencing of Human Chromosome 16
              Unpublished
REFERENCE     2 (bases 1 to 151049)
              DOE Joint Genome Institute.
              Direct Submission
              Submitted (14-DEC-1999) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
              3 (bases 1 to 151049)
              DOE Joint Genome Institute.
              Direct Submission
              Submitted (02-NOV-2001) Production Sequencing Facility, DOE Joint
              Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
              On Nov 2, 2001 this sequence version replaced gi:9795566.
              Sequence Quality Assessment:
              This entry has been annotated with sequence quality
              estimates computed by the Phrap assembly program.
              All manually edited bases have been reduced to quality zero.
              Quality levels above 40 are expected to have less than
              1 error in 10,000 bp.
```

Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

-----  
Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.

All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

-----  
FEATURES  
source Location/Qualifiers

1. .151049  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-80F22"  
BASE COUNT 46827 a 30783 c 30749 g 42690 t  
ORIGIN

Query Match 99.88; Score 999.4; DB 9; Length 151049;

Best Local Similarity 99.9%; Pred. No. 1.1e-299;

Matches 1000; Conservative 0; Mismatches 1; Indels 0; Gaps 0;

QY 1 GGAGATGATAAACCGTGTAGTGCCTCAAGTTGTGTGGACCATGGAATGGGAGACTG 60

Db 137896 GGAGATGATAAACCGTGTAGTGCCTCAAGTTGTGTGGACCATGGAATGGGAGACTG 137955

QY 61 GAGGATACATGGATGCCCAACTACAGGCCCGCTCCAGTATGAGCCATGAGCCAGTT 120

Db 137956 GAGGATACATGGATGCCCAACTACAGGCCCGCTCCAGTATGAGCCATGAGCCAGTT 138015

QY 121 GAATCTGAATGTGAAGTGAATGAAGACCGCAGAGAGTACACGTCAACCCCTCAT 180

Db 138016 GAATCTGAATGTGAAGTGAATGAAGACCGCAGAGAGTACACGTCAACCCCTCAT 138075

QY 181 AACATGGGTGAGATCAAGAAACACACACAGAGCTGAGAACTGGTGTAGTCCAGGG 240

Db 138076 AACATGGGTGAGATCAAGAAACACACACAGAGCTGAGAACTGGTGTAGTCCAGGG 138135

QY 241 TCAGGCAAAAACCCCTGACTCCATGTTTATGGCCATGCTAGCTGTAATATCTGTGAGT 300

Db 138136 TCAGGCAAAAACCCCTGACTCCATGTTTATGGCCATGCTAGCTGTAATATCTGTGAGT 138195

QY 301 ATGATTTTCTGTGAGAGCAAAACATATATGGGCATATTTTCTTAACCCACCGGTAGT 360

Db 138196 ATGATTTTCTGTGAGAGCAAAACATATATGGGCATATTTTCTTAACCCACCGGTAGT 138255

QY 361 GTGATCATCTCTGAAGCAGACATCTCTCTGAGATATATCATGATCAAGGAGCATCAGTA 420

Db 138256 GTGATCATCTCTGAAGCAGACATCTCTCTGAGATATATCATGATCAAGGAGCATCAGTA 138315

QY 421 CCAGGACCTCTAACTCCCGCTGACACAGAGCAATTAGACTCTCTATAACAATGGTATCAAT 480

Db 138316 CCAGGACCTCTAACTCCCGCTGACACAGAGCAATTAGACTCTCTATAACAATGGTATCAAT 138375

QY 481 TATACCACTCATTTGGAGGACTTCTTTTATGTGTACCCAGGATACATGCTCAACTGTC 540

Db 138376 TATACCACTCATTTGGAGGACTTCTTTTATGTGTACCCAGGATACATGCTCAACTGTC 138435

QY 541 AGTTGCCCTTGAGTTGATCCCAAGCATGGTTGAGTTACCATATAAAATATCTACCTA 600

Db 138436 AGTTGCCCTTGAGTTGATCCCAAGCATGGTTGAGTTACCATATAAAATATCTACCTA 138495

QY 601 TTAGACCTTAGCTTTATTAATATTACTGTGTAGTTACTATCACTCTCTGCCCCCATCAC 660

Db 138496 TTAGACCTTAGCTTTATTAATATTACTGTGTAGTTACTATCACTCTCTGCCCCCATCAC 138555

QY 661 CCAAAATGTACTGATTTACAGAAATGGCTCCCTTTGATTAATCTACCCCTCTCTTGG 720

|||||

Db 138556 CCAAAATGTACTGATTTATACAGAAATGGCTCCCTTTGATAAATCTTCAACCCCTCTTGG 138615

QY 721 GCCCACTGCTTGGCCCTTAGCTAGACAAATAGTCCATGTTAATGGGAGACATTATTGAC 780

Db 138616 GCCCACTGCTTGGCCCTTAGCTAGACAAATAGTCCATGTTAATGGGAGACATTATTGAC 138675

QY 781 TGGGTCCCTCTGTGTCATTAAAGATGGGAGAGATGAGAATCAGACACATGGCATAAACTT 840

Db 138676 TGGGTCCCTCTGTGTCATTAAAGATGGGAGAGATGAGAATCAGACACATGGCATAAACTT 138735

QY 841 CACTGGCACTGGTGGCGAACTTTAAACATCTTTCACATTCACACACATGGGATCAATCC 900

Db 138736 CACTGGCACTGGTGGCGAACTTTAAACATCTTTCACATTCACACACATGGGATCAATCC 138795

QY 901 CAATCTGCCATGCAACTTGTTCATGCAATGGAACGGCTTTAGCCACCTTGCCTCAATGG 960

Db 138796 CAATCTGCCATGCAACTTGTTCATGCAATGGAACGGCTTTAGCCACCTTGCCTCAATGG 138855

QY 961 CATTATCAAGGAAAGAGAGGTCCCAATTCAGGAGTCTATGTG 1001

Db 138856 CATTATCAAGGAAAGAGAGGTCCCAATTCAGGAGTCTATGTG 138896

RESULT 3

AC113268/c

LOCUS

AC113268

DEFINITION

AC113268

ACCESSION

VERSION

AC113268.8

GI:22095124

HTG.

KEYWORDS

SOURCE

olive baboon.

ORGANISM

Papio cynocephalus anubis

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Cercopithecoidea;

Cercopithecinae; Papio.

REFERENCE

1 (bases 1 to 191014)

Zhou,L., Fu,Y., Wu,J., Eichler,E. and Roe,B.A.

Papio anubis BAC Clone rp41-22m16

Unpublished

REFERENCE

2 (bases 1 to 191014)

Zhou,L., Fu,Y., Wu,J., Eichler,E. and Roe,B.A.

Direct Submission

Submitted (06-AUG-2002)

Department Of Chemistry And Biochemistry,

The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,

OK 73019, USA

REFERENCE

3 (bases 1 to 191014)

Zhou,L., Fu,Y., Wu,J., Eichler,E. and Roe,B.A.

Direct Submission

Submitted (06-AUG-2002)

Department Of Chemistry And Biochemistry,

The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,

OK 73019, USA

REFERENCE

4 (bases 1 to 191014)

Zhou,L., Fu,Y., Wu,J., Eichler,E. and Roe,B.A.

Direct Submission

Submitted (14-AUG-2002)

Department Of Chemistry And Biochemistry,

The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,

OK 73019, USA

COMMENT

On Aug 4, 2002 this sequence version replaced gi:21104942.

----- Genome Center

Center: Department Of Chemistry And Biochemistry

The University Of Oklahoma

Center code:UOKNOR

FEATURES

source

Location/Qualifiers

1. .191014

/organism="Papio cynocephalus anubis"

/sub\_species="anubis"

/db\_xref="taxon:9555"

/clone="rp41-22m16"

/clone\_lib="RPCI - 41 Male (Olive) Baboon BAC Library"

54882 a 40134 c 39560 g 56438 t

BASE COUNT

ORIGIN

```
Query Match      83.6%; Score 836.8; DB 9; Length 191014;
Best Local Similarity 91.5%; Pred. No. 5.6e-249;
Matches 922; Conservative 0; Mismatches 77; Indels 9; Gaps 3;

QY 1 GGAGATGATAAACCGCTGAGTGCCTCAAGTTCTGTGCGACCATGGAATGGGAGACTG 60
Db 109942 GGAGATGATAAACCATGTAGTGCCTCAAGGTGTGTGACCATGGAATGGGAGACTG 109883

QY 61 GAGGATACATGATGCCAACTACAGGCCAGCTCCTCAGTATGAGCCATGAGCCAGTT 120
Db 109882 GAGGATACATGATGCCAACTACAGGCTTGCTCCTCAGTATGAGCCATGAGCCAGTT 109823

QY 121 GAATCTGAATGTGAAGATGGAATGAAGACCGAGAGTCACTAGCGTCAACCCCTCAT 180
Db 109822 GAATCTGAATGTGAAGATGGAATGAAGACTGACCAGAGCCATGTACATCAACCCCTCAT 109763

QY 181 RACATGGGGTCAGATCAAGAAACACACAGGAGCTGAGAACTGGTGTAGTCCAGGG 240
Db 109762 AACGTGGGTACGGTTCAAGAAACACACAGGAGCTGAGAACTGGTGGAGTCCAGGG 109703

QY 241 TCAGGCAAAAACCCCTGACTCCATGTTTATGGCCATGCTAGCTGTAATATCCTGTGCAGT 300
Db 109702 TCAGGCAAAAACCCCTGACTCCATGTTTATGGCCATGCTAGCTGTAATATCCTGTGCAGT 109643

QY 301 ATGATTTTCTGTGCAGAAAGCAAAACATATTGGGCATATTTTCTCAACCCACCGGTAGT 360
Db 109642 ATGATTTTCTGTGCAGAAAGCAAAACATATTGAGTGTATGTTCCCAACCCACCGGTAGT 109583

QY 361 GTGATCATACTCTGAAGCAGCACTCCTCTGAGATATATCATGATCAAGGAGCATCAGTA 420
Db 109582 GTGACCATACTCTGAAGTGCACACTCCTCTGAGGATATATCATGATCAAGGAGTATAGTA 109523

QY 421 CCAGGACCTCTAACPCCCCTGACACAGAGCAATTAGACTCTCTAFAACAATGGTATCAAT 480
Db 109522 CCAGGACCTCTAACPCCCCTGACACAGAGCAATTAGACTCTCTAFAACAATGGTATCAAT 109463

QY 481 TATACACTCCATTTGAGGGACTCCTTTATGTGTACCCAGGATACATATGCTCAACTGC 540
Db 109462 TATATCACTCCATTTGAGGGACTCCTTTATGTGTACATCAACAGGATGATCGTCAACTGC 109404

QY 541 AGTTCCTTGCAAGTTGATCCCAAGCATGTTGAGTTACCA-----TAAATAAATTAT 593
Db 109403 AGTTCCTTGCAAGTTGATCCCAAGCATGTTGAGTTACCAAAAAAAAAAAAAAATTAT 109344

QY 594 GTACCTATTAGACCTTAGCTTTATTAATATTACTTGTGTAGTACTAATCACTCCTGGCC 653
Db 109343 GTACCTATTAGACCTTAGCTTTATTAATATTATTGTTGTAGTACTAATCACTCCTGGCC 109284

QY 654 CCATCAACCAAAATGCTAGTATTACAGAAATGGGCTCCTTTGATTAATTCACCCGCC 713
Db 109283 CCATCAACCAAAATGCTAGTATTACAGAAATGGGCTCCTTTGATTAATTCACCA-CCGCC 109225

QY 714 TCCTTGGGCCCACTCTCTTGGCCCTTAGCTAGACAATAGTCCATGTTAATGGGAGACAT 773
Db 109224 TCCTTGGGCCCACTCTCTTGGCCCTTAGCTAGACAATAGTCCATGTTAATGGGAGACAT 109165

QY 774 TATTGACTGGGTCCTGTGTGTCATTAAGATGGGAGAGATGAGAAATCAGACACATGGCA 833
Db 109164 TATTGACTGGGTCCTGTGTGTCATTTAGATAGGAGAGATGAGAAATCAGACCTCATGGCA 109105

QY 834 TAAACTTCACCTGGCACTGGTGGCGAAACTTTAAACATCTCTTCACATTCACACACTGGGAT 893
Db 109104 TAAACTTCACCTGGCACTGGTGGCGAAACTTTAAACATCTCTTCACATTCACACACTGGGAT 109045

QY 894 TCAATCCCAATCTGCCATCAACTTGTGGCATGGAAGGGCTTTAGCCACCTTTGCC 953
Db 109044 TCAATCCCAATCTGCCATCAACTTGTGGCATGGAAGGGCTTTAGCCACCTTTGCC 109885

QY 954 TCAATGGCATTTAAGGAAAGAGAGGTCCTCAATTCAGGAGTCTATGTG 1001
Db 109884 TCAATGGCATTTAAGGAAAGAGAGGTCCTCAATTCAGGAGTCTATGTG 108937
```



Db 88837 AACATGGGACAGATCAAGAAAAACACACAGAGAAAGTTGAAAACTGCTGGAGCGCCAGGG 88896  
QY 241 TCAGGCAAAACCCCTGACTCCATGTTATGATGCTAGCTAGTGTAAATATCTGTGCACT 300  
Db 88897 TCAGGCAAAACCCAGACTCCATGTTATGATGCTAGCTAGTGTAAATATCTGTGAGT 88956  
QY 301 ATGATTTTCTGTGAGAGCAAAACATATTTGGGCATATTTTCCCTAACCCACCCTGAGT 360  
Db 88957 ATGTTTCTGCTGAGAGCAAAACATATTTGGGCATATTTTCCCTAACCCACTGGTAGT 89016  
QY 361 GTGATCATACTCTGAGAGCACTCTCTGAGATATATCATGATCAAGGACATCAGTA 420  
Db 89017 GTGACATACTCTGAGAGCACTCTCTGAGATATATCATGATCAAGGACATGAGCA 89076  
QY 421 CCAGGACCTCTAACCTCCCTGACACAGAGCAATTAGACTCTCATAAACAAATGGTATCAAT 480  
Db 89077 CCAGGACCTCTAACCTCCCTGACACAGAGCAATTAGACTCTCATGACAAATGGTATCAAT 89136  
QY 481 TATACACTCCATTTGGAGGACTTCTTTATGTGTCACCCAGGATACATGTCCTCAACTGC 540  
Db 89137 TATACACTCCATTTGGAGGACTTCTTTATGCGTCACCCAGGATACCTCAACTGC 89196  
QY 541 AGTTGCTTGGAGTTGATCCCAAGCATGTTGAGTTACCATATAAAAAAATTTATGTACCTA 600  
Db 89197 AGTTGCTTGGCAATCAATCCCAAGCATGCTGAGTTACCATGGAATATTTATGTACCTA 89256  
QY 601 TTAGACCTTAGCTTATTAATATTAATCTGTGTAGTTACTTAATCACTCTCTGGCCCATCAC 660  
Db 89257 TTAGGCTTAGCTTATTAATATTAATCTGTGTAGTTACTTAATCACTCTCTGGCCCATCAC 89316  
QY 661 CCAATTTGACTGATATACAGAAATGGGCTCCTTTGATATTAATCTCACCCCTCTCTTGG 720  
Db 89317 CCAATTTGATTTGATATACAGAAATGGGCTCCTTTGATATTAATCTCACCCCTCTCTTGG 89376  
QY 721 GCCCACTGTTCGCCCTTACTAGACATAATAGTCCATGTTAATGGGACATATTATGAC 780  
Db 89377 GCCCACTGTTCGCCCTTACTAGACATAATAGTCCATGTTAATGGGACATATTATGAC 89436  
QY 781 TGGGTCCCTGTGTCATTAGATGGGAGATGAGATCAGACACATGCGCATAAACTT 840  
Db 89437 TGGGTCCCTGTGTCATTAGATGGGAGATGAGATCAGACACATGCGCATAACTT 89496  
QY 841 CACTGGCACTGTGCGCAAACTTTAACATCTCTTCACTACATCACTGGGATTCATGCC 900  
Db 89497 CACTGGCACTGTGCGCAAACTTTAACATCTCTTCACTACATCACTGGGATTCATGCC 89556  
QY 901 CAATCTGCATCACTTGTGCGATGAGCGGCTTTAGCCCACTTTGCTTCAATGG 960  
Db 89557 CAATCTGCATCACTTGTGCGATGAGCGGCTTTAGCCCACTTTGCTTCAATGG 89616  
QY 961 CATTATCAAGGAAAGAGGTCCTCAAT 987  
Db 89617 CATTATCAAGGAAAGAGGTCCTCAAT 89643

RESULT 5 AC003682 153875 bp DNA linear PRI 29-JUN-1998  
LOCUS AC003682 Homo sapiens chromosome 19, overlapping cosmids F18547, F11133,  
DEFINITION R27945, R28830 and R32804, complete sequence.  
ACCESSION AC003682  
VERSION AC003682.1 GI:3264845  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 153875)  
REFERENCE Lamerdin, J.E., McCreedy, P.M., Skowronski, E., Adamson, A.W.,  
AUTHORS Burkhardt-Schultz, K., Gordon, L., Kyle, A., Ramirez, M., Stillwagen, S.,  
Phan, H., Velasco, W., Do, L., Regala, W., Terry, A., Barnes, J.,  
Danganan, B., Poundstone, P., Christensen, M., Georgescu, A., Avila, J.,  
Liu, S., Attix, C., Andreise, T., Frankheim, M., Amico-Keller, G.,

Cosfield, J., Duarte, S., Lucas, S., Bruce, R., Thomas, P., Quan, G.,  
Kronmiller, B., Arellano, A., Montgomery, M., Ow, D., Nolan, M.,  
Trong, S., Kobayashi, A., Olsen, A.S. and Carrano, A.V.  
Sequence analysis of a 2 Mb region containing a zinc finger (ZNF)  
gene cluster in 19q13.4  
Unpublished  
2 (bases 1 to 153875)  
Lamerdin, J.E.  
Direct Submission  
Submitted (16-DEC-1997) Human Genome Center, Lawrence Livermore  
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA  
3 (bases 1 to 153875)  
Lamerdin, J.E.  
Direct Submission  
Submitted (29-JUN-1998) Joint Genome Institute, Lawrence Livermore  
National Laboratory, 7000 East Ave., Livermore, CA 94551, USA  
On Jun 29, 1998 this sequence version replaced gi:2689440.  
Map and sequence oriented from q centromere to telomere.  
This accession derived from cosmid F18547 from bases 1 to 37,736,  
cosmid F11133 from bases 36,861 to 40,261, cosmid R27945 from bases  
40,194 to 77,593, cosmid R28830 from bases 74,683 to 118,033, and  
cosmid R32804 from bases 114,809 to 153,875. This accession  
overlaps cosmid F25419 (AC003005) to the left from bases 1 to  
5,157, and overlaps cosmid F20191 (AC004017) from bases 150,049 to  
153,875. No sequence discrepancies were identified in any of the  
overlapping regions between cosmids.

Additional chr 19 map and sequence information are available at:  
http://www-bio.lnl.gov/bbrp/genome/genome.html.

FEATURES  
source  
Location/Qualifiers  
1..153875  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/map="19q13.4 between D19S773 to ZNF134"  
/clone="Splice of sequence from cosmids  
F18547-F11133-R27945-R28830-R32804"  
/cell\_lines="UV5HL9-5B for F library clones, 5HL2-B for R"  
/note="Cosmid libraries LL19NC02 and LL19NC03 were  
constructed at LNL from flow-sorted chromosomes from  
human-hamster hybrids UV5HL9-5B and 5HL2-B, respectively,"  
which carry chromosome 19 as their only human chromosome."  
complement(1..62)  
/rpt\_family="Alu"  
101..248  
/rpt\_family="L1"  
247..808  
/rpt\_family="L1MB8"  
813..1114  
/rpt\_family="TIGGER2"  
complement(1115..1325)  
/rpt\_family="AluSc"  
1329..1752  
/rpt\_family="LTR3"  
2245..4769  
/rpt\_family="HERVK"  
6786..7215  
/rpt\_family="LTR3"  
complement(7217..7296)  
/rpt\_family="AluY"  
7518..7794  
/rpt\_family="AluJo"  
complement(7795..7919)  
/rpt\_family="AluJo/FLAM"  
7923..8719  
/rpt\_family="TIGGER2"  
complement(8723..9023)  
/rpt\_family="AluSq"  
9024..9489  
/rpt\_family="TIGGER2"  
9498..9529  
/rpt\_family="MER28"  
9530..9607

```
repeat_region /rpt_family="L1ME1"
repeat_region /rpt_family="AluY"
repeat_region complement(10335..10637)
repeat_region /rpt_family="AluSg"
repeat_region 11021..11319
repeat_region /rpt_family="AluY"
repeat_region complement(12033..12161)
repeat_region /rpt_family="GGA)n"
repeat_region complement(13427..13598)
repeat_region /rpt_family="AluY"
repeat_region complement(13612..13912)
repeat_region /rpt_family="AluSx"
repeat_region 13941..14242
repeat_region /rpt_family="AluSp"
repeat_region 15000..15288
repeat_region /rpt_family="AluSx"
repeat_region complement(15458..15699)
repeat_region /rpt_family="MER4B"
repeat_region complement(16629..17123)
repeat_region /rpt_family="LINE2"
repeat_region complement(17670..17917)
repeat_region /rpt_family="MER7A"
repeat_region 17941..18164
repeat_region /rpt_family="MER44A"
repeat_region 18323..18616
repeat_region /rpt_family="AluY"
repeat_region 18617..18639
repeat_region /rpt_family="TAAA)n"
repeat_region 18640..18801
repeat_region /rpt_family="MER44C"
repeat_region complement(18842..19438)
repeat_region /rpt_family="MER21B"
repeat_region complement(19578..19716)
repeat_region /rpt_family="MIR"
misc_feature 19782..19820
frame: 1, quality: excellent, score: 76.000"
repeat_region /note="predicted exon, program: graill2exons_human_1.3,
repeat_region 20111..20668
repeat_region /rpt_family="MER44C"
repeat_region complement(20669..20930)
repeat_region /rpt_family="L1"
repeat_region 21073..21180
repeat_region /rpt_family="MER54"
repeat_region complement(21667..21830)
repeat_region /rpt_family="AluSg"
repeat_region complement(21832..22641)
CDS join(23850..23978,25861..27639)
similar to zinc finger protein ZNF132 (U09411)[Homo sapiens]
/codon_start=1
/product="F18547_1"
/protein_id="AAC24605.1"
/db_xref="GI:2689441"
/translation="MOGHVTFDIAYVFSQEWGLDRAORCLYHDVLMENFSLMASL
PSPKVCILYFCIEMWGLGCHIEAEAPSEQTLISAQVSQARTPKLGPSPINRAHSC
MCILVMKDLILSEHQGLPWQKFTSVASGNWFSFGLNQQHQNQDSGEKHRRKES
SALLNCSLIPSLNDFCKDVKDFTPLGLLQHTTHSROEYAHRSRETFQQRXYK
CEQVFNKRVHQRVHTGKAYKREYKSLNSKYLFEVHQRTHNAEPYVNCIG
KSFJHQTGLVGHQRIHTRERSYVCIEGKSLSSKYLSEVHQRTHNGEKPYVCNVCQK
SPRIKOTPVGHQRIHTGERPYVCMCEGKSFHSYDRIHORVHTGEGAYOCSECGKS
FIYKOSLDLHRIHTGERPYECKGKAFIHKRLLEHQRHTGKPYVCLICKSPFI
RSSDMRHRHRTGERPYECKGKAFISKOTLLKHKIHTRERPYECKGKGFYLE
VKLLOHQRHTREOLCEGKGFVSHQKRLLEHQRVHTGKPECEGKGFRRHS
LIHQKVSHERPYNACEKAFIYKKNLHQRHTGKPEYCKGKGFRRHS
RHQKVHTEEP"
misc_feature 23856..23978
frame: 2, quality: good, score: 72.000"
repeat_region /note="predicted exon, program: graill2exons_human_1.3,
repeat_region 24902..25017
repeat_region /rpt_family="MER47B"
```

```
repeat_region 25019..25213
repeat_region /rpt_family="MER47B"
repeat_region 25519..25814
repeat_region /rpt_family="AluJb"
misc_feature 25855..27639
frame: 0, quality: excellent, score: 84.000"
misc_feature 28460..28498
frame: 2, quality: good, score: 53.000"
repeat_region 29190..29486
repeat_region /rpt_family="AluSx"
repeat_region 29656..29954
repeat_region /rpt_family="AluSx"
repeat_region 31008..31058
repeat_region /rpt_family="L1M4"
repeat_region 31339..31368
repeat_region /rpt_family="AT_Rich"
repeat_region complement(31425..31613)
repeat_region /rpt_family="MLT1B"
misc_feature complement(31821..31892)
frame: 0, quality: excellent, score: 83.000"
repeat_region 32367..32667
repeat_region /rpt_family="AluSg"
repeat_region 33552..33847
repeat_region /rpt_family="AluSx"
repeat_region 34313..34436
repeat_region /rpt_family="FLAM_C"
repeat_region 34443..34735
CDS /rpt_family="AluJo"
complement(join(35687..36705,44943..45069))
frame: 0, quality: excellent, score: 83.000"
Most similar to PID1d1024590 (AB007873) KIAA0413 [Homo sapiens]
/codon_start=1
/product="R27945_1"
/protein_id="AAC24606.1"
Query Match 72.68; Score 726.8; DB 9; Length 153875;
Best Local Similarity 86.18; Pred. No. 1.1e-214; Indels 32; Gaps 4;
Matches 863; Conservative 0; Mismatches 107;
Qy 1 GGAGATGATAAACCGTGTGAGTGCCTCAAGTTGTGCGACCATGGAATGGGAGACTG 60
Db 4765 GGAGATGACAAATCGTGTGGTGCCTCAAGTTGTGACCATGGAACGGGAGACTG 4824
Qy 61 GAGGATACATGGATCCCAACTACAGGCCAGCTCCTCCAGTATGAGCCATGAGCCAGTT 120
Db 4825 GAGAAATCCATGGATCTCAACTGTGGGCCAGCTCCTCCAGTATGAGCCATGAGCTAGTT 4884
Qy 121 GAATCTGAATGAGATGGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 180
Db 4885 GAATCTGAATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 4944
Qy 181 AACATGGGTGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGATGAGAT 240
Db 4945 AAATGGGAGAGCTCAGAAACACACAGGAGAGCTGAGAACTGCTGGAG----- 4996
Qy 241 TCAGCAAAAACCCCTGACTCCATCTTTATGGCCATGAGTCTGCTGCTGCTGCTGCTGCTGCT 300
Db 4997 -----CAAAACCCCTGACTCCATCTTTATGGCCATGAGTCTGCTGCTGCTGCTGCTGCT 5050
Qy 301 ATGATTTTCTGTCAGAGCAAAACATATGAGGATATTTCCCTAACCCACCGGTAGT 360
Db 5051 ATGTTTCTGTCAGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGG 5110
Qy 361 GTGATC-ATATCTCTGAAGCAGCAGCTCCCTCCCTGAGATATATCATGATCATGAGGATCAGT 419
Db 5111 AGCACCCTGACTCGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 5170
Qy 420 ACCAGGAGCTTAACTCCCTCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 479
Db 420 ACCAGGAGCTTAACTCCCTCCCTGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAGGAG 479
```

Db 5171 ACCAGGACCCCTAACTCCCTGACAGAGATGATTAAGATTCTCAGAACAAATGGTATCA 5230  
Qy 480 TTATACCACTCCATGGAGGACTTCCTTTATGTGTCCACCAGGATACATGCTCAACTG 539  
Db 5231 TTATACGACCAATGGAGGACTT-CITTTATGTGTCCACCAGGATGCATCACTCAACCG 5289  
Qy 540 CAGTTGCTTGCAGTTTGATCCCAAGCATGTTGAGTTACCAATAAAAAATATGTACCT 599  
Db 5290 CAGTTGCTTGCATTAATCAGTCCCAAGCATGCTTGAAGTTACCAATGAAAAATATGTACCT 5349  
Qy 600 ATTAGACCTTACCTTTATTAATATTACTGTGTAGTTACTTAATCACTCTGCGCCCATCA 659  
Db 5350 ATTAGGCTTACCTTATTAATATTACTGTGTAGTTACTTAATCACTCTGCGCCCATCA 5409  
Qy 660 CCCAAATGTACTGATTATACAGAAATGGGCTCCCTTTTGATTAATCTCACCCCTCCCTTG 719  
Db 5410 CCCAAATGTATTGATTGTACAGAAATGGGCTCCCTTTTGATTAATCTCACCCCTGCTCTG 5469  
Qy 720 GGGCCACTGCTTGGCCCTTAGCTAGACAAATAGTCCATGTTAATGGGAGACATATTGA 779  
Db 5470 GACTCAGTGTCTTGGCCCTTAGCTAACAACAGTCCATGTTAATGGGAGACATATTGA 5529  
Qy 780 CTGGGCTCCCTGTGCTAATGAATGGGAGATGAGAAATCAGACACATGCGATAACT 839  
Db 5530 CTGGGCTCCCTGTGCTAATGATGGGAGATGAGAAATCAGACCTTCAATGGCATAAT 5589  
Qy 840 TCAGTGCACCTGGGCGAACTTTAATCATCTTCACTTCAACACACATGGGATTCATC 899  
Db 5590 TTGCTGCACCTGGGCGAACTTTGAACATCTCTCGCTAC-----ATC 5633  
Qy 900 CCAATCTGCCATGCACTTGTGGCATGGAACGGGCTTTAGCCACCTTTGCTCCTCAATG 959  
Db 5634 CCAATCTGCTACACAACCTTGTGGCAGAAATGGGCTTTAGCCACCTTTGCTCCTCAATG 5693  
Qy 960 GCATTATCAAGGAAGAGAGTCCAAATCAGGAGTCTATGTG 1001  
Db 5694 GCATTATCAAGGAAGAGAGTCCAAATTCAGGAGTCAATATG 5735

AC010642 45383 bp DNA linear PRI 26-AUG-2000  
LOCUS Homo sapiens chromosome 19 clone LUNLR-245B6, complete sequence.  
DEFINITION AC010642  
ACCESSION AC010642  
VERSION AC010642.5 GI:9929687  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 45383)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 45383)  
DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (16-SEP-1999) Production Sequencing Facility, DOE Joint  
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
AUTHORS 3 (bases 1 to 45383)  
DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (26-AUG-2000) DOE Joint Genome Institute, 2800 Mitchell  
Drive, Walnut Creek, CA 94598, USA  
COMMENT On Aug 26, 2000 this sequence version replaced gi:7711487.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www-shgc.stanford.edu  
Quality: Phrap Quality >=40 99.5% of Sequence;  
Estimated Total Number of Errors is 0.3.  
SVS Content:  
SHGC-13293 G13764

WI-3037 G04140.  
FEATURES Location/Qualifiers  
source 1. 45383  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19"  
BASE COUNT 12088 a 11041 c 11063 g 11191 t  
ORIGIN  
Query Match 64.9%; Score 649.2; DB 9; Length 45383;  
Best Local Similarity 81.0%; Pred. No. 1.4e-190;  
Matches 81; Conservative 0; Mismatches 168; Indels 23; Gaps 4;  
Qy 1 GGAGATGATAAAGCGGTGTGAGTGCCTCAAGTTGTGTGCGACCATGGAATGGGAGACTG 60  
Db 8105 GGAGATGGACAACCGTGTGGTGCCTCCAGGTGTGTGACCATGGAACGGGAGCCG 8164  
Qy 61 GAGGATACATGGATCCCAACTACAGGCCAGCTCCTCCAGTATGAGCCATGAGCCAGTT 120  
Db 8165 GAGGATCCATGTATTCAACCGTGGGCTGTACCTCCAGTAGGACCATGAGCCAGCG 8224  
Qy 121 GAATCTGAATGTGAAGATGGAATGAAGACCGACGAGAGTCACTGAGTCAACCCCTCAT 180  
Db 8225 GAATCTGAATGCAAGACAGAAAGGCCGCCAGGAGTCAATGACATCCAAACCCCAT 8284  
Qy 181 AACATGGGCTCAGATCAAGAAAACACACAGAAAGCTGAGAAACTGGTGTAGTGCCAGGG 240  
Db 8285 AACATGGGACAGATCAAGAAAACGACACAGAAAGCTGAGAAACTACTGGAGCCGAGGG 8344  
Qy 241 TCAGGCAAAAACCCCTGACTCCATGTTTATGAGCCATGCTAGCTGTAATATCCTGTGCAGT 300  
Db 8345 TCAGGCAAAAACCCCTGACTCCATGTTTCTTGGCCATGCTAGTGTCTGTGCGGT 8404  
Qy 301 ATGATTTTCTGTGCGAAGCAAAAACATATTTGGGCATATTTTCTTAACCCACCGGTAGT 360  
Db 8405 ATGTTTCCCTGTGTAGAGGCAAAAACATATTTGGGCATATTTTCTTAACCCACCTGGTAGT 8464  
Qy 361 GTGA-TCATACTCTGAAGCAGCACTCCTCTGAGATATATCATGATCAAGGAGCATCAGT 419  
Db 8465 ACAGTGGTACTCTGGAGCGACACTCCTCTGAGATATATATGATCAGGAGGCGTGGCG 8524  
Qy 420 ACCAGGACCTTAATCCCTGACACAGAGCAATTAAGACTCTCATAAACATGTTATCAA 479  
Db 8525 ACCAGGACCCCTAACTCCGTCTGACACAGAG-----CAATAGCATCAA 8567  
Qy 480 TTATACCACTCCATGGAGGAGCTTCCTTTATGTGTCCACCAGGATACATGCTCAACTG 539  
Db 8568 TTACACCGGCCCATTTGGAAGGACTCCCTTTATGTATCACCATGATACGTCACCTCAACTG 8627  
Qy 540 CAGTTGCTTGCAGTTTGTATCCCAAGCATGGTTGAGTTTACCATAAAAAAATATGTACCT 599  
Db 8628 CAGTGTCTGCATCCAACTCAGGATGTTTAAAGTCAACCTGGAANAATATGTATTT 8687  
Qy 600 ATTAGACCTTACCTTTATTAATATTACTGTGTAGTTACTTAATCACTCTGCGCCCATCA 659  
Db 8688 ATTAGGCTCAGCTTTATTAATGTTACTGGCGTGTCCACCAATCAGTCTAGTACCCCATCA 8747  
Qy 560 CCCAAATGTACTGATTATACAGAAATGGGCTCCCTTTTGATTAATTTCTACACCCCTCTTG 719  
Db 8748 CTCAAATGTATGATTATATATGGAATGGGCTCCCTTTTGATATTTCTCACCCCTCTCTTG 8807  
Qy 720 GGGCCACTGCTTGGCCCTTAGCTAGACAAATAGTTCATGTTAATGGGAGACATATTGA 779  
Db 8808 GACCCAGTGTCTTGGCCCTTGGCCGACACACAGTCTATGTTAATGGGAAC---ATTGA 8864  
Qy 780 CTGGGCTCCCTGTGCTAATGAATGGGAGAGATGAGAATCAGACCATGAGCAATAACT 839  
Db 8865 CTGGGCTCCCGTGGCCATTTAGAGGGGAGAGATGAAAATCAGACCTCTTGTGGCAAAACT 8924  
Qy 840 TCAGTGCACCTGGGCGAAACTTTAATCATCTTCTAGTCTCAACACACTGGGATTCAT- 898  
Db 8925 TCAGTGCACCTGGTGGGAGCTTCAACACCTCTTCAATCATCATCACATGCGATTTCAGT 8984

```
Qy 899 -CCCAATCTGCCATGCAACTGCTTGGCATGGAAGGGCTTTAGCCCACTTTGCTGCTCA 957
    ||| || ||||| ||| ||| ||||| ||| ||| |||||
Db 8985 CCCAGTTGGCTGCACAACCTTGCTGGTATAGGACAGGCTTTAAACCCGCCCTTACCTCAG 9044
    ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 958 TGGCATTTATCAAGGAAGAGAGGTCCTCAATTCAGGAGTCTATGTG 1001
    ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 9045 TGACGTCTATGGAACAGAGGCCCAATTCAGGAATCGATATG 9088
    ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

RESULT 7
AC020915 160852 bp DNA linear PRI 31-MAY-2001
DEFINITION Homo sapiens chromosome 19 clone CTD-3138B18, complete sequence.
ACCESSION AC020915
VERSION AC020915.6 GI:14269683
KEYWORDS HTG.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Unpublished
REFERENCE
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (12-JAN-2000) Production Sequencing Facility, DOE Joint
Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.
TITLE Direct Submission
JOURNAL Submitted (31-MAY-2001) DOE Joint Genome Institute, 2800 Mitchell
Drive, Walnut Creek, CA 94598, USA
COMMENT On May 31, 2001 this sequence version replaced gi:13752690.
Draft Sequence Produced by DOE Joint Genome Institute
www.jgi.doe.gov
Finishing Completed at Stanford Human Genome Center
www.shgc.stanford.edu
Quality: Phrap Quality >=40 99.9% of Sequence;
Estimated Total Number of Errors is 0.1.
STS Content:
SHGC-9787 GL1320.

FEATURES
    source
        1. 160852
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="19"
            /clone="CTD-3138B18"
BASE COUNT 42283 a 37137 c 39005 g 42427 t
ORIGIN

Query Match 64.9%; Score 649.2; DB 9; Length 160852;
Best Local Similarity 81.0%; Pred. No. 1.7e-190;
Matches 813; Conservative 0; Mismatches 168; Indels 23; Gaps 4;

Qy 1 GGAGATGATAAACCGTGTGAGTGCCTCAAGTCTGTGCGACCATGGAATGGAGACTG 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 159499 GGAGATGGACAACCGTGTGGGTGCCCTCCAGGTGTGTGTGACCATGGAACGGGAGACCG 159558
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 61 GAGGATACATGGATCCCAACTACAGGCCCGACGCTCCTCCAGTATGAGGCATGAGCCAGTT 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 159559 GAGGATCCATGGTATTCAACCGTGGCGCTGTGTACCTCCAGTACGACCATGAGCCAGCG 159618
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 121 GAATCTGAATGTGAAGATGAATGAAGACCGACGAGAGTCACTAGCTCAACCCCTAT 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 159619 GAATCTGAATGCAAGACAGACAGAGCGGCCGACCGAGTCAACAATGACATCCAAACCCCAT 159678
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Qy 181 AACATGGGTCAATCAAGAAACACACACAGCAAGCTGAGAACTGGTGTAGTGCAGGG 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 159679 AACATGGGACAGATCAAGAAACACACACAGCAAGCTGAGAACTGAGTGGAGCGCCAGGG 159738
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```

University School of Medicine, 4444 Forest Park Parkway, St. Louis,  
MO 63108, USA  
On May 19, 2000 this sequence version replaced gi:7219897.

## COMMENT

----- Genome Center -----  
Center: Washington University Genome Sequencing Center  
Center code: WUGSC  
Web site: http://genome.wustl.edu/gsc/index.shtml  
----- Project Information -----  
Center project name: H.NH0706G10  
----- Summary Statistics -----  
Sequencing vector: M13; 100%  
Sequencing vector: plasmid; 0%  
Chemistry: Dye-terminator ET; 100% of reads  
Assembly program: Phrap; version 0.990319  
Consensus quality: 208624 bases at least Q40  
Consensus quality: 209946 bases at least Q30  
Consensus quality: 210614 bases at least Q20  
Insert size: 203000; agarose-fp  
Insert size: 212442; sum-of-contigs  
Quality coverage: 6.38 in Q20 bases; agarose-fp  
Quality coverage: 6.07 in Q20 bases; sum-of-contigs  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 17 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

```
1 1486: contig of 1486 bp in length
* 1487 1588: gap of unknown length
* 1587 3987: contig of 2401 bp in length
* 3988 4087: gap of unknown length
* 4088 7555: contig of 3468 bp in length
* 7556 7655: gap of unknown length
* 7656 9830: contig of 2175 bp in length
* 9831 9931: gap of unknown length
* 9931 12684: contig of 2754 bp in length
* 12685 12785: gap of unknown length
* 12785 18591: contig of 5807 bp in length
* 18592 18691: gap of unknown length
* 18692 25938: contig of 7147 bp in length
* 25939 30672: gap of unknown length
* 30673 30773: gap of unknown length
* 30773 36865: contig of 6092 bp in length
* 36865 44900: gap of unknown length
* 44901 45000: contig of 7936 bp in length
* 45001 58755: gap of unknown length
* 58756 58856: gap of unknown length
* 58856 72460: contig of 13605 bp in length
* 72461 72560: gap of unknown length
* 72561 92043: contig of 19483 bp in length
* 92044 92144: gap of unknown length
* 92144 113070: contig of 20926 bp in length
* 113070 113170: gap of unknown length
* 113170 137968: contig of 24799 bp in length
* 137969 138069: gap of unknown length
* 138069 175883: contig of 37814 bp in length
* 175883 175983: gap of unknown length
* 175983 214042: contig of 38060 bp in length.
```

FEATURES  
source

```
Location/Qualifiers
1..214042
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="19"
/clone="RP11-706G10"
1..1486
/notes="assembly_name:Contig7"
```

```
misc_feature 1587..3987
/notes="assembly_name:Contig8"
misc_feature 4088..7555
/notes="assembly_name:Contig9"
misc_feature 7656..9830
/notes="assembly_name:Contig10"
misc_feature 9931..12684
/notes="assembly_name:Contig11"
clone_end:T7
vector_side:left"
vector_side:right"
misc_feature 12785..18591
/notes="assembly_name:Contig12"
misc_feature 18692..25838
/notes="assembly_name:Contig13"
misc_feature 25939..30672
/notes="assembly_name:Contig14"
misc_feature 30773..36864
/notes="assembly_name:Contig15"
misc_feature 36965..44900
/notes="assembly_name:Contig16"
misc_feature 45001..58755
/notes="assembly_name:Contig17"
misc_feature 58856..72460
/notes="assembly_name:Contig18"
vector_side:right"
misc_feature 72561..92043
/notes="assembly_name:Contig19"
misc_feature 92144..113069
/notes="assembly_name:Contig20"
misc_feature 113170..137968
/notes="assembly_name:Contig21"
misc_feature 138069..175882
/notes="assembly_name:Contig22"
misc_feature 175983..214042
/notes="assembly_name:Contig23"
BASE COUNT 54973 a 51053 c 49621 g 56788 t 1607 others
ORIGIN
```

Query Match 64.9%; Score 649.2; DB 2; Length 214042;  
Best Local Similarity 81.0%; Pred. No. 1.7e-190;  
Matches 813; Conservative 0; Mismatches 168; Indels 23; Gaps 4;

```
Qy 1 GGAGATGATTAACCGTGTGAGTCCCTCAAGTTGTGTGGACCATGAATGGGAGACTG 60
Db 43581 GGAGATGATTAACCGTGTGAGTCCCTCAAGTTGTGTGGACCATGAATGGGAGACTG 43640
Qy 61 GAGGATACATGGATCCCAACTACAGCCGCCAGCTCCCTCCAGTATGAGCCATGAGCCAGTT 120
Db 43641 GAGGATACATGGATCCCAACTACAGCCGCCAGCTCCCTCCAGTATGAGCCATGAGCCAGTT 120
Qy 121 GAATCTGAATGTGAAGATGAAGAACCGACGAGAGTACACAGTACAGTCAACCCATCAT 180
Db 43701 GAATCTGAATGTGAAGATGAAGAACCGACGAGAGTACACAGTACAGTCAACCCATCAT 43760
Qy 181 AACATGGGTCAGATCAAGAAAACACACAGAGAGTGAAGAACTGGTGTAGTSCCAGGG 240
Db 43761 AACATGGGTCAGATCAAGAAAACACACAGAGAGTGAAGAACTGGTGTAGTSCCAGGG 43820
Qy 241 TCAGGCAAAAACCCCTGACTCCATGTTTATGGCCATGCTAGCTGTAATATCTCTGTGCAGT 300
Db 43821 TCAGGCAAAAACCCCTGACTCCATGTTTATGGCCATGCTAGCTGTAATATCTCTGTGCAGT 300
Qy 301 ATGATTTTCTGTGCAGAACAAAACATATTTGGGCATATTTTCTTAACCCACGGTAGT 360
Db 43881 ATGATTTTCTGTGCAGAACAAAACATATTTGGGCATATTTTCTTAACCCACGGTAGT 43940
Qy 361 GTGA-TCATACTCTGAAGCAGCACTCTCTCTGAGATATATCATCATCAAGAGCATCAGT 419
Db 43941 ACAGTCTGTACTCTGGAGCGACACTCTCTCTGAGATATATATATCATCAGGAGCGTGGC 4400
Qy 420 ACCAGGACCTCTTAACCTCCCTGACACAGACCAATTAGACTCTCATTAACAAATGGTATCAA 479
```

Db 44001 ACCAGAGCCCTTAACCTCCGCTGCACAGAG-----CAATAGCATCAA 44043  
Qy 480 TTATACCACTCCATGGAGGACTTCCTTTATGTGTCCACGAGATACATTCCTCAACTG 539  
Db 44044 TTACACCGGCCCATTTGGAGGACTCCCTTTATGTATCACCATGGATAGTCACTCACTG 44103  
Qy 540 CAGTTGCCCTTGCACTTTGATCCCAAGCAGTGTGTAGTTACCATAAAAAAATATGTACTT 599  
Db 44104 CAGCTGTCTTGCACTTCACTCAGGCATCGTAAATGACCATGGAAAAAATATGTATTT 44163  
Qy 600 ATTACACCTTAGCTTTATTAATATTAATTAATTAATTAATTAATTAATTAATTAATTA 659  
Db 44164 ATTAGGCTCAGCTTTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 44223  
Qy 660 CCCAAATGTACTGATTAATACAGAAATGGCTCCCTTTGATTAATTAATTAATTAATTAAT 719  
Db 44224 CTCAAATGTATGATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 44283  
Qy 720 GCCCAGCTGCTTGGCCCTTGTAGCTAGACATAATAGTCCATGTTAATGGGAGACATTTATGA 779  
Db 44284 GACCCAGTGTCTTGGCCCTTGGCCAGACAAACAGTCTATGTTAATGGGAAC---ATTGA 44340  
Qy 780 CTGGGGTCCCTGTGCTATTAGATGGGAGAGATGAGATCAGACCATGGCATAACT 839  
Db 44341 CTGGGGTCCCGTGGCCATTTAGAGGGGAGAGATGAAAATCAGACCTCTTGCCACAACT 44400  
Qy 840 TCACCTGGCACTGGTGGCGAACTTTAAACATCTTCACTTCAACACACTGGGATTCAT- 898  
Db 44401 TCACCTGGCACTGGTGGCGAACTTTCAACACCTTTCATCATCATCATCGGATTCAGTC 44460  
Qy 899 -CCCAATCTGCCATCAACTTCTTGGCATGGAACGGCTTTAGCCCACTTTGGCTCAA 957  
Db 44461 CCCAGTGGCTGCACAACTTCTTGGTATAGGACAGGCTTTAAACCCGCTTACCTCAG 44520  
Qy 958 TGGCATTATCAGGAAGAGAGAGTCCATTCAGGAGTCTATGTG 1001  
Db 44521 TCACGCTATCTAGGAACAGAGGCGCCAAATTCAGGAATCGATATG 44564

RESULT 9  
AC01589  
DEFINITION Homo sapiens clone RP11-44L20, WORKING DRAFT SEQUENCE, 32 unordered pieces.  
AC01589  
VERSION AC01589.4 GI:10047664  
KEYWORDS HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE Homo sapiens  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 158802)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
TITLE Homo sapiens chromosome, clone RP11-44L20  
JOURNAL Unpublished  
AUTHORS 2 (bases 1 to 158802)  
REFERENCE 1 (bases 1 to 158802)  
AUTHORS Birren,B., Linton,L., Nusbaum,C., Lander,E., Allen,N., Anderson,M., Baldwin,J., Barna,N., Beckerly,R., Boguslavskiy,L., Boukhalter,B., Brown,A., Castle,A., Colangelo,M., Collins,S., Collumore,A., Cooke,P., Dearrellano,K., Dewar,K., Domino,M., Doneilan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C., Funke,R., Gage,D., Galagan,J., Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L., Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J., Lehoczy,J., Lieu,C., Locke,K., Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J., Meldrum,J., Morrow,J., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P., Peterson,K., Pollard,V., Riley,R., Roy,A., Santos,R., Severy,P., Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Tirrell,A., Vassiliev,H., Vo,A., Wheeler,J., Wu,X., Wyman,D., Ye,W.J., Zimmer,A. and Zody,M.  
DIRECT SUBMISSION  
TITLE Direct Submission  
JOURNAL Submitted (17-NOV-1999) Whitehead Institute/MIT Center for Genome Research, 320 Charles Street, Cambridge, MA 02141, USA

## COMMENT

On Sep 10, 2000 this sequence version replaced gi:7107718.  
All repeats were identified using RepeatMasker:  
Snit, A.P.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence\_submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L1096  
Center clone name: 44\_L20  
----- Summary Statistics  
Sequencing vector: M13; M7815; 100% of reads  
Chemistry: Dye-primer-amersham; 6% of reads  
Chemistry: Dye-terminator Big Dye; 94% of reads  
Assembly program: Phrap; version 0.960731  
Consensus quality: 139045 bases at least Q40  
Consensus quality: 146903 bases at least Q30  
Consensus quality: 150628 bases at least Q20  
Insert size: 170000; agarose-fp  
Quality coverage: 3.2 in Q20 bases; agarose-fp  
Quality coverage: 3.5 in Q20 base.  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 32 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
\* 1 5488: contig of 5488 bp in length  
\* 5489 5588: gap of 100 bp  
\* 5589 7161: contig of 1573 bp in length  
\* 7162 7261: gap of 100 bp  
\* 7262 8657: contig of 1396 bp in length  
\* 8658 8757: gap of 100 bp  
\* 8758 10159: contig of 1402 bp in length  
\* 10160 10259: gap of 100 bp  
\* 10260 12259: contig of 2000 bp in length  
\* 12260 12359: gap of 100 bp  
\* 12360 14250: contig of 1891 bp in length  
\* 14251 14350: gap of 100 bp  
\* 14351 16194: contig of 1844 bp in length  
\* 16195 16294: gap of 100 bp  
\* 16295 17825: contig of 1531 bp in length  
\* 17826 17925: gap of 100 bp  
\* 17926 19224: contig of 1299 bp in length  
\* 19225 19324: gap of 100 bp  
\* 19325 21412: contig of 2088 bp in length  
\* 21413 21512: gap of 100 bp  
\* 21513 23847: contig of 2335 bp in length  
\* 23848 23947: gap of 100 bp  
\* 23948 25938: contig of 1991 bp in length  
\* 25939 26038: gap of 100 bp  
\* 26039 28260: contig of 2222 bp in length  
\* 28261 28360: gap of 100 bp  
\* 28361 31500: contig of 3140 bp in length  
\* 31501 31600: gap of 100 bp  
\* 31601 34208: contig of 2608 bp in length  
\* 34209 34308: gap of 100 bp  
\* 34309 37683: contig of 3375 bp in length  
\* 37684 37783: gap of 100 bp  
\* 37784 42304: contig of 4521 bp in length  
\* 42305 42404: gap of 100 bp  
\* 42405 46675: contig of 4271 bp in length  
\* 46676 46775: gap of 100 bp  
\* 46776 60170: contig of 13395 bp in length  
\* 60171 60270: gap of 100 bp  
\* 60271 64808: contig of 4536 bp in length  
\* 64807 64906: gap of 100 bp  
\* 64907 69745: contig of 4839 bp in length

```
* 69746 69845: gap of 100 bp
* 69846 75307: contig of 5462 bp in length
* 75308 75407: gap of 100 bp
* 75408 81675: contig of 6268 bp in length
* 81676 81775: gap of 100 bp
* 81776 87139: contig of 5364 bp in length
* 87140 87239: gap of 100 bp
* 87240 93647: contig of 6408 bp in length
* 93648 93747: gap of 100 bp
* 93748 103371: contig of 9624 bp in length
* 103372 103471: gap of 100 bp
* 103472 112431: contig of 8960 bp in length
* 112432 112531: gap of 100 bp
* 112532 121885: contig of 9354 bp in length
* 121886 121985: gap of 100 bp
* 121986 130194: contig of 8209 bp in length
* 130195 130294: gap of 100 bp
* 130295 142016: contig of 11722 bp in length
* 142017 142116: gap of 100 bp
* 142117 155225: contig of 13109 bp in length
* 155226 155325: gap of 100 bp
* 155326 158802: contig of 3477 bp in length.
FEATURES
    source
    1..158802
        /organism="Homo sapiens"
        /db_xref="taxon:9606"
        /clone="RP11-44L20"
        /clone_lib="RPCI-11 Human Male BAC"
        1..5488
            /note="assembly_fragment
            clone_end:SP6
            vector_side:left"
            5589..7161
                /note="assembly_fragment"
            7262..8657
                /note="assembly_fragment"
            8758..10159
                /note="assembly_fragment"
            10260..12259
                /note="assembly_fragment"
            12360..14250
                /note="assembly_fragment"
            14351..16194
                /note="assembly_fragment"
            16295..17825
                /note="assembly_fragment"
            17926..19224
                /note="assembly_fragment"
            19325..21412
                /note="assembly_fragment"
            21513..23847
                /note="assembly_fragment"
            23948..25938
                /note="assembly_fragment"
            26039..28260
                /note="assembly_fragment"
            28361..31500
                /note="assembly_fragment"
            31601..34208
                /note="assembly_fragment"
            34309..37683
                /note="assembly_fragment"
            37784..42304
                /note="assembly_fragment"
            42405..46675
                /note="assembly_fragment"
            46776..60170
                /note="assembly_fragment"
            60271..64806
                /note="assembly_fragment"
            64907..69745
                /note="assembly_fragment"
            69846..75307
```

```
misc_feature
75408..81675
/note="assembly_fragment"
misc_feature
81776..87139
/note="assembly_fragment"
misc_feature
87240..93647
/note="assembly_fragment"
misc_feature
93748..103371
/note="assembly_fragment"
misc_feature
103472..112431
/note="assembly_fragment"
misc_feature
112532..121885
/note="assembly_fragment"
misc_feature
121986..130194
/note="assembly_fragment"
misc_feature
130295..142016
/note="assembly_fragment"
misc_feature
142117..155225
/note="assembly_fragment"
misc_feature
155326..158802
/note="assembly_fragment
clone_end:T7

Query Match      62.7%; Score 627.4; DB 2; Length 158802;
Best Local Similarity 79.1%; Pred. No.1e-183;
Matches 795; Conservative 0; Mismatches 181; Indels 29; Gaps 3;

Qy 1 GGAGATGGATAAACCGTGTGAGTGCCTCAAGTTGTGTGCGCACCATGGAATGGGAGACTG 60
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 110087 GGAGATGGACAACCATGTGGTGGCTGCCCTCCAGGTGTGCGCACCATGAATGGGAGACTA 110146
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 61 GAGGATACATGGATCCCAACTACAGGCCCCAGCTCCTCCAGTATGAGCATGAGCCAGCTT 120
||| ||| ||| ||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 110147 GCAGGACCCAGTTGGCCAAACCGTGGACCTGGTCCCTCCAGTAGAGTCTGAGCCAGCT 110206
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 121 GAATCTGAATGTGAAGATGGAAATGAAGACCGCAGAGTACACTGACGTCAAGTCAACCTCAT 180
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 110207 GAATTTGAATGCAAGATGGAGCAAGGCCGACTGGAGTCACTGACATCCACCCCAT 110266
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 181 AACATGGGTCAGATCAAGAAAACCAACACAGAGACTGAGAAACTGGTGTAGTGCCAGGG 240
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 110267 AACATGGGACAGATCAAGAAAACCAAGAGAGCTGAGAAAATGCTGGAGCG----- 110320
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 241 TCAGGCAAAAACCCCTGACTCCATGCTTATGGCCATGCTAGCTGTATATATCCTGTGCAGT 300
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 110321 -----TCCATGTGCTTGGCCATGTTACGAGTAGTGTCTGTGAGGT 110361
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 301 ATGATTTTCTGTGCAGAAGCAAAACATATTGGGCATATTTTCTAACCCACCGGTAGT 360
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 110362 ATGTTTCCCTGTGTAGAGGCCAAAACATCTGGGCATATTTCTTAATCAACCTGTAGT 110421
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 361 GTGATC-ATACTCTGAAGCAGCACTCCTCCTGAGATATATCATCATCAAGGAGCATCAGT 419
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 110422 GTGACCTGTACTCTGGAGTGACATTCCTCTCTGAGGTATATCATCATCAGGAGCGTGGC 110481
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 420 ACCAGGACCTCTAACTCCCTGACAGAGCAATAGACTCTCATACAAATGCTATCAA 479
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 110482 TCTAGGCCCCCTAACTCCCTGATATAGACATTAACATCTCACACATATAGATCAG 110541
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 480 TTATACCACCTCCATTGGAGGACTTCCTTTATGTGTCAACCAGGATACATTTGCTCAACTG 539
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 110542 TTATACGCCCCCATTTGGAGGACTCCCTTTATGTCATCACCAGAAGATATGTCACTCAACCA 110601
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 540 CAGTTGCTTGCAGTTTGATCCCAAGCATGGTTGAGTTACCATAAAAAAATATGTACTCT 599
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 110602 CAGTTGCTTGCAGTCCCAACCCAGGACTGGTTGAGTCACCATGAAAAGTATGTATTT 110661
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 600 ATTAGACTTAGCTTTTATTAATATTTACTTGTGTAGTTACTAATCACTCTCGCCCCCATCA 659
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 110662 ATTAGGCTTAGCTCTTATTAATACTAGTGTCTCATCAACCACTCCCGCCCCCATCA 110721
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||

Qy 660 CCCAAATTGTACTGATTATACAGAAATGGCTCCCTTTTGATTAATTTCT---CACCCCCCTCC 716
||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
```



|            |  |                 |
|------------|--|-----------------|
| Db 110722  | CAAAATTGTTATTGATTATACAGATGGCTCCCTTTGATAGTTCTACCACCCCTCC            | 110781          |
| Qy 717     | TTGGGCCCACTGCTGTGGCCCTTAGCTAGACAATPAGTCCATGTTAAATGGGAGACATPAT      | 776             |
| Db 110782  | TTGGACCAATGCTCTGTGCTCTTTAGCTTAGACAACAGTCTATGCTAAATGGGAGACATTGG     | 110841          |
| Qy 777     | TGACTGGGGTCCCTGTGCTCATTAGATGGGAGAGATGAGATCAGACCATGGCATAA           | 836             |
| Db 110842  | TGACTGGGGACCCCGTGTATTATAGATGGAAGAGATGAGATCAGACCTCATGGSCAAA         | 110901          |
| Qy 837     | ACTTCACCTGGCACTGGTGGCGAACTTTAACATCTCTTCACTTCAACACACTGGGATTCA       | 896             |
| Db 110902  | ACTTCGCTGGCACTGGTGGCGACACTTCAACATCTCTTCACTACTACATCAGACTGGGATTCA    | 110961          |
| Qy 897     | ATCCCAATCTGCCATCAACTTGTCTGGCATGGAACGGGCTTTAGCCACACTTTGCCCTCA       | 956             |
| Db 110962  | ATCCCACTGGCTGTGGCTTGTCTGGCACAGACGAGCCTTAGCTCGCCTTGGCCTCA           | 111021          |
| Qy 957     | ATGGCATTTATCAAGAAAGAGAGTCCAAATTCAGGAGTCTATGTG                      | 1001            |
| Db 111022  | GTGGCAGCATCAAGGAAAGAGGGGCCCAATTCAGGAGTCGATATG                      | 111066          |
| RESULT 10  |  |                 |
| AC010328   | 196686 bp DNA linear   | PRI 13-JUL-2002 |
| LOCUS      | Homo sapiens chromosome 19 clone CITB-EL_2620122, complete         |                 |
| DEFINITION | sequence.  |                 |
| ACCESSION  | AC010328   |                 |
| VERSION    | AC010328.4   | GI:7381733      |
| KEYWORDS   | htg.   |                 |
| SOURCE     | human.   |                 |
| ORGANISM   | Homo sapiens   |                 |
| REFERENCE  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  |                 |
| AUTHORS    | Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.         |                 |
| TITLE      | DOE Joint Genome Institute and Stanford Human Genome Center.       |                 |
| JOURNAL    | 1 (bases 1 to 196686)  |                 |
| REFERENCE  | DOE Joint Genome Institute.  |                 |
| AUTHORS    | 2 (bases 1 to 196686)  |                 |
| TITLE      | Direct Submission  |                 |
| JOURNAL    | Submitted (15-SEP-1999) Production Sequencing Facility, DOE Joint  |                 |
| REFERENCE  | Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA |                 |
| AUTHORS    | 3 (bases 1 to 196686)  |                 |
| TITLE      | DOE Joint Genome Institute and Stanford Human Genome Center.       |                 |
| JOURNAL    | Submitted (01-APR-2000) DOE Joint Genome Institute, 2800 Mitchell  |                 |
| REFERENCE  | Drive, Walnut Creek, CA 94598, USA                                 |                 |
| AUTHORS    | 4 (bases 1 to 196686)  |                 |
| TITLE      | DOE Joint Genome Institute and Stanford Human Genome Center.       |                 |
| JOURNAL    | Submitted (18-APR-2000) DOE Joint Genome Institute, 2800 Mitchell  |                 |
| REFERENCE  | Drive, Walnut Creek, CA 94598, USA                                 |                 |
| AUTHORS    | 5 (bases 1 to 196686)  |                 |
| TITLE      | DOE Joint Genome Institute and Stanford Human Genome Center.       |                 |
| JOURNAL    | Submitted (13-JUL-2002) DOE Joint Genome Institute, 2800 Mitchell  |                 |
| REFERENCE  | Drive, Walnut Creek, CA 94598, USA                                 |                 |
| AUTHORS    | On Apr. 1, 2000 this sequence version replaced gi:6997076.         |                 |
| TITLE      | Draft Sequence Produced by DOE Joint Genome Institute              |                 |
| JOURNAL    | www.jgi.doe.gov  |                 |
| COMMENT    | Finishing completed at Stanford Human Genome Center                |                 |
| FEATURES   | Quality: Phrap Quality >=40 99.5% of Sequence;                     |                 |
| source     | Estimated Total Number of Errors is 0.9.                           |                 |
|            | STS Content:   |                 |
|            | SHGC-32227 G31446.   |                 |
|            | Location/Qualifiers  |                 |
|            | 1..196686  |                 |
|            | /organism="Homo sapiens"   |                 |
|            | /db_xref="taxon:9606"  |                 |
|            | /chromosome="19"   |                 |

|                           |   |        |
|---------------------------|---|--------|
| BASE COUNT                | 50946 a 43545 c 45142 g 57053 t                                 |        |
| ORIGIN                    |   |        |
| Query Match               | 62.5%; Score 625.8; DB 9; Length 196686;                        |        |
| Best Local Similarity     | 79.0%; Pred. No. 3.4e-183;                                      |        |
| Matches 794; Conservative | 0; Mismatches 182; Indels 29; Gaps 3;                           |        |
| Qy 1                      | GGAGATGGATAAACCGTGTGAGTGGCCCTCAAGTTGTGTGCGACCATGGAATGGGAGACTG   | 60     |
| Db 159354                 | GGAGATGGACAAACCATGCTGGGTGCCCTCCAGGTGTGTGCGACCATGAATGGGAGACTA    | 159413 |
| Qy 61                     | GAGGATACATGGATCCCACTACAGGCCCACTCCCTCCAGCATGAGCCATGAGCCAGTT      | 120    |
| Db 159414                 | GCAGGACCAACCGTTGGCCCAACCGTGGACCTGCTCCCTCCAGTAGGAGTCGTGAGCCAGCT  | 159473 |
| Qy 121                    | GAATCTGAATCTGAAGATGGAATGAAGACCGACGAGAGTCACTGACGTCAACCCCTCAT     | 180    |
| Db 159474                 | GAATTTGAATGCAAGATGGAGCAAGGCCCACTGGAGTCACTGACATCCACCCCAT         | 159533 |
| Qy 181                    | AACATGGGGTCCAGATCAAGAAAAACACACAGAACTGAGAACTGGTGTAGTGCCAGGG      | 240    |
| Db 159534                 | AACATGGGGCAGATCAAGAAAAACAAAGCAGGAAGCTGAGAAAAATGCTGGA-----       | 159584 |
| Qy 241                    | TCAGGCAAAAACCCCTGACTCCATGTTTATGGCCATGCTAGCTGTAATATCTGTGCACT     | 300    |
| Db 159585                 | -----GCATCCATGTGCTTGGCCATGTTAGCAGTAGTGTCTGTGAGGT                | 159628 |
| Qy 301                    | ATGATTTTCTGTGCAGAAAGCAAAAACATATTGGGCATATTTTCTTAACCCACCGGTAGT    | 360    |
| Db 159629                 | ATGTTTCCCTGTGTAGAGGCCAAACATATCTGGCATATGTTCCCTAATACCCCTGTAGT     | 159688 |
| Qy 361                    | GTGATC-ATACTCTGAAGCAGCACTCCTCTCGAGATATATCATCATCAAGGAGCATCAT     | 419    |
| Db 159689                 | GTGACCTGTACTCTGGAGTGACATTTCCCTCTGAGGTATATCATCATCAGGAGCGTGGC     | 159748 |
| Qy 420                    | ACCAGGACCTTAACCTCCCTCGACAGCAAGCAATTAGACTCTCATACAATGGTATCAA      | 479    |
| Db 159749                 | TCTAGGGCCCTTAACCTCCCTCGATATAGCAATTAACACTCTCACAACAATAGATCAG      | 159808 |
| Qy 480                    | TTATACCACTCCATTTGAGGAGCACTTCTCTTATGTCTACCCAGGATACATCTGCTCAACTG  | 539    |
| Db 159809                 | TTATACCCGCCCATTTGAGGAGCACTCCCTTTATGCATCACCAGAATATGTCACTCAACA    | 159868 |
| Qy 540                    | CAGTTGCCCTTGACGTTTGATCCCAAGCATGGTTGAGTTTACCATAAAAAATATATGACT    | 599    |
| Db 159869                 | CAGTTGTCTTGACAGTCCAACCCCGGCACTGGTTGAGTCACCATGGAAAAGTTATGTATTT   | 159928 |
| Qy 600                    | ATTAGACCTTAGCTTTATTAATATTAATTAATTAATTAATTAATTAATTAATTAATTAAT    | 659    |
| Db 159929                 | ATTAGGCCCTTAGCTCTATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA   | 159988 |
| Qy 660                    | CCCAATTTGATGATTTATACAGATGGCTCCCTTTGATTAATTTCT---CACCCCTCC       | 716    |
| Db 159989                 | CCAAATTTGATGATTTATACAGATGGCTCCCTTTGATTAATTTCT---CACCCCTCC       | 160048 |
| Qy 717                    | TTGGGCCCACTGCTGTGGCCCTTAGCTAGACAAATPAGTCCATGTTAAATGGGAGACATPAT  | 776    |
| Db 160049                 | TTGGACCAATGCTCTGTGCTCTTTAGCTAGACAACAGTCTATGCTAAATGGGAGACATTGG   | 160108 |
| Qy 777                    | TGACTGGGGTCCCTGTGCTCATTAGATGGGAGAGATGAGATCAGACCATGGCATAA        | 836    |
| Db 160109                 | TGACTGGGGACCCCGTGGTTAATTAGATGGAAGAGATGAGATCAGACCTCATGGCAAA      | 160168 |
| Qy 837                    | ACTTCACCTGGCACTGGTGGCGAACTTTAACATCTCTTCACTTCAACACACTGGGATTCA    | 896    |
| Db 160169                 | ACTTCGCTGGCACTGGTGGCGACCTTCAACATCTCTTCACTACTACATCAGACTGGGATTCA  | 160228 |
| Qy 897                    | ATCCCAATCTGCCATGCAACTGCTGTGGCATGGAACGGGCTTTAGCCCACTTTGCCCTCA    | 956    |
| Db 160229                 | ATCCCACTGCGGCTGTGCGGCTTGTCTTGGCAGACAGCAGGCTTTAGCTCGCCTTTGCCCTCA | 160288 |
| Qy 957                    | ATGGCATTTATCAAGGAAAGAGAGGTCCCAATTCAGGAGTCTATGTG                 | 1001   |





```
repeat_region /rpt_family="Alu"
7359..7481
repeat_region /rpt_family="MaLR"
7482..7577
repeat_region /rpt_family="HY3"
7583..7627
repeat_region /rpt_family="(TAAA)n"
7628..7663
repeat_region /rpt_family="MaLR"
7777..8072
repeat_region /rpt_family="Alu"
8144..8252
repeat_region /rpt_family="L1"
8250..8265
misc_feature /note="match to EST BM015942 (NID:g16530296),"
8253..8549
repeat_region /rpt_family="Alu"
8550..9057
repeat_region /rpt_family="L1"
9058..9181
repeat_region /rpt_family="Alu"
9112..9405
misc_feature /note="match to EST BG719345 (NID:g13998532),"
9182..9279
repeat_region /rpt_family="L1"
9280..9604
repeat_region /rpt_family="Alu"
9605..9693
repeat_region /rpt_family="L1"
9694..10001
repeat_region /rpt_family="Alu"
10002..10356
repeat_region /rpt_family="L1"
10464..10832
misc_feature /note="similar to Sus scrofa EST AW479624 (NID:g7049730),"
11011..11093
repeat_region /rpt_family="MIR"
11312..11638
repeat_region /rpt_family="Alu"
11320..11335
misc_feature /note="match to EST AW022240 (NID:g5875770) df34g09.yl"
11322..11335
misc_feature /note="match to EST AI207473 (NID:g6361481)"
11845..11851
misc_feature /note="match to EST BE294168 (NID:g9177699)"
12037..12101
repeat_region /rpt_family="Alu"
12558..12623
misc_feature /note="match to EST AA010360 (NID:g1471396) z108f11.sl"
12382..13251
misc_feature /note="CpG island (%GC=65.6, o/e=0.80, #CpGs=77)"
12488..12751
misc_feature /note="match to EST BG719345 (NID:g13998532)"
12759..13292
misc_feature /note="match to EST BG720169 (NID:g13999356)"
13168
misc_feature /note="match to EST AA010359 (NID:g1471395) z108f11.rl"
13213..13292
misc_feature /note="match to EST AA010359 (NID:g1471395) z108f11.rl"
13327..13330
```

Query Match 61.4%; Score 615; DB 9; Length 124102;  
Best Local Similarity 86.7%; Pred. No. 7.4e-180;  
Matches 703; Conservative 0; Mismatches 100; Indels 8; Gaps 2;

```
Qy 191 CAGATCAAGAAACACACACAGAGCTGAGAACTGGTGTAGTCCAGGGTCAGGCAAAA 250
Db 107348 CAGATCAAGAAACACACACAGAGATGAGAACTGCTGGAGCATCAGGGTCAGGCAAAA 107289
Qy 251 ACCCTGACTCCATGTTTATGCCATGCTAGCTGTGAATATCCTGTGTCAGTATGATTTTC 310
Db 107288 ACCCTGACTCCATGTTTATGCCATGCTAGCTGTGAATATCCTGTGTCAGTATGATTTTC 107229
```

```
Qy 311 TGTGAGAGCAAAAACATATTTGGGCATATTTTCTTAACCCACCGGTAGTGTGATCATAC 370
Db 107228 TGTGAGAGGCAAAAACATATTTGGGCATATTTTCTTAACCCACCGGTAGTATGA ----- 107175
Qy 371 TCTGAAGCAGCACTCCCTCCTGAGATATATCATGATCAAGAGGAGCATCAGTACCAGGACCTC 430
Db 107174 -CCGGGGTGACACTCCTCCTGAGATATATCATGATCAAGAGGAGCATGCGGACGACCT 107116
Qy 431 TAATCCCTCCCTGACACAGAGCAATTAGACTCTCATAAATGGTATCAATATATACCACTC 490
Db 107115 TAATCCCTCCCTGACACAGAGCAATTAGACTCTCAGAACAATGGTATCAATATATACCACTC 107056
Qy 491 CATTGGAGGACTTCCTTTATGTGTCAACCCAGGATACATTTGCTCAACTGCAAGTTCGCTTG 550
Db 107055 CATTGGAGGGCTTTCTTTATGTCATCACCCAGGATACATTTGCTCAACTGCAAGTTCGCTTG 106996
Qy 551 CAGTTTGTATCCCAAGCATGGTTGAGTTACCATAAAAAATTTATGTACCTATTAGACCTTA 610
Db 106995 CAATTCAATCCCGAGCATGGTTGAGTTACCATGGAATAATTTACGTACCTATTGGGCCCTTA 106936
Qy 611 GCTTTATTAAATATTACTTGTGTAGTTACTAATCACTCCTCGGCCCATCACCCAAATTTGTA 670
Db 106935 GTTCTACTAATATTACTGGTGTAGTTACTAATCACTCCTCGGCCCATCACCCAAATTTGTA 106876
Qy 671 CTGATTATACAGATGGGCTCCCTTTGATAAATCTCACCCCTCCCTTGGGCCCACTGTC 730
Db 106875 TTAATTATATAGATGGGCTCCCTTTGATAAATCTCAACCCCTCCCTTGGGCCCTGTC 106816
Qy 731 TTGGCCCTTTAGCTAGACAAATAGTCCATGTTTAATGGGAGACATTTATGACTTGGGTCCT 790
Db 106815 TTAGTCCCTTAGCTAGACAAACAGTCCATGCTTAATGGGAGACATTTATGACTTGGGTCCT 106756
Qy 791 GTGTGCTAATTAAGATGGGAGATGAGATCAGACATGAGACATGGCATAAATCTTCACTGGCACT 850
Db 106755 ATGTCATTTGTTGGGAGACATGAGCATCAGACCTCATGTCATTAATTTGCTTGGCACT 106696
Qy 851 GGTGGCAAACTTTAACAATCTTTCACATTCACACACATGGGATTCATCAATTCGCA 910
Db 106695 GGTGGCAAACTTTAACAATCTC-CCACTACATCAGATTCGCAATTCATCTCAATCTGCTG 106637
Qy 911 TGCAACTGCTTGGCATGGGAACGGCTTTAGCCACCTTTGCCCAATGGCATATATCAAG 970
Db 106636 CCCAAATGCTTGACACAGCATGCTTTAGCTACCTTTGCCCAATGGCATATATCAAG 106577
Qy 971 GAAAGAGAGTCCCAATTCAGGAGTCTATGTG 1001
Db 106576 GAAAGAGAGTCCCAATTCAGGAGTCAATATG 106546
```

## RESULT 12

AC069481

LOCUS

DEFINITION Homo sapiens chromosome 5 clone RP11-153J19 map 5, WORKING DRAFT  
SEQUENCE, 28 unordered pieces.

AC069481

VERSION

KEYWORDS

SOURCE

ORGANISM

human.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE

AUTHORS

TITLE

JOURNAL

REFERENCE

AUTHORS

2 (bases 1 to 158302)

Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,

Anderson,S., Baldwin,J., Barna,N., Bastien,V., Beda,F.,

Boguslavsky,L., Boukhgalter,B., Brown,A., Burkett,G.,

Campomayor,A., Castle,A., Choepel,Y., Colangelo,M., Collins,S.,

Collamore,A., Cooke,P., DeArelano,K., Dewar,K., Diaz,J.S.,

Dodge,S., Domino,M., Doyle,M., Ferreira,P., Fitzhugh,W., Gage,D.,

Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,



```
misc_feature 25827..27337
              /note="assembly_fragment"
misc_feature 27438..29802
              /note="assembly_fragment"
misc_feature 29903..32473
              /note="assembly_fragment"
misc_feature 32574..35307
              /note="assembly_fragment"
misc_feature 35408..38588
              /note="assembly_fragment"
misc_feature 38689..41368
              /note="assembly_fragment"
misc_feature 41469..44056
              /note="assembly_fragment"
misc_feature 44157..48600
              /note="assembly_fragment"
misc_feature 48701..52600
              /note="assembly_fragment"
misc_feature 52701..58166
              /note="assembly_fragment"
misc_feature 58267..62266
              /note="assembly_fragment"
misc_feature 62367..65972
              /note="assembly_fragment"
misc_feature 66073..72761
              /note="assembly_fragment"
misc_feature 72862..79897
              /note="assembly_fragment"
misc_feature 7998..86476
              /note="assembly_fragment"
misc_feature 86577..94565
              /note="assembly_fragment"

Query Match      61.4%; Score 615; DB 2; Length 158302;
Best Local Similarity 86.7%; Pred. No. 7.7e-180;
Matches 703; Conservative 0; Mismatches 100; Indels 8; Gaps 2;

QY 191 CAGATCAAGAAAACACACAGCAAGCTGAGAACTGGTGTAGTGCAGGGTCAGGCAAAA 250
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 78925 CAGATCAAGAAAACACACAGAAAGATGAGAACTGCTGGAGCATCAGGTCAGGCAAAA 78984

QY 251 ACCCTGACTCCATGTTTATGGCCATGCTAGCTGTAATATCTGTGCATGATATTTTC 310
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 78985 ACCCTGACTCCATGTTTATGGCCATGCTAGCTGTAATATCTGTGTGATGTTTCCC 79044

QY 311 TGTGCAGAGCAAAAACATATTTGGGCATATTTCTTAACCCACCGTAGTGTGATCATAC 370
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 79045 TGTGCAGAGCAAAAACATATTTGGGCTTATGTTCTTAACCCCGGTAGTATGA----- 79098

QY 371 TCTGAAGCAGCACTCTCTCTGAGATATATCATGATCAAGGAGCATCAGTACCAGGACCTC 430
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 79099 -CGGGGTGACACTCTCTCTGAGATATATCATGATCAAGGAGCATGGCACCAGGACCTC 79157

QY 431 TAACTCCCTGACAGAGCAATTAAGACTCTCAATCAATGATGATCAATTAATACACTC 490
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 79158 TAACTCCCTGACAGAGCAATTAAGACTCTCAGAACAAATGATCAATTAATACACTC 79217

QY 491 CATTGGAGGACTCTCTTTATGTGTCACCCAGGATACATGCTCAACTCAGTTCCTTG 550
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 79218 CATTGGAGGCTCTCTTTATGTCATCCAGGATACATGCTCAACTCAGTTCCTTG 79277

QY 551 CAGTTTATCCCAAGCATGTTGAGTTAGCATAAAAAATATGTAACCTATTAGACTTA 610
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 79278 CAATTCAATCCAGGATGTTGAGTTACCATGGAATTAATGATGATGTTGGGCTTA 79337

QY 611 GCTTTTAAATTAATTAATGTTAGTAACTCACTCTGCGCCCATCAATCAATTAATGTA 670
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 79338 GTTCTACTAATAATTAATGTTAGTAACTCACTCTGCGCCCATCAATCAATTAATGTA 79397

QY 671 CTGATTATACAGAAATGGCTCCCTTTGATAATTTCTCAACCCCTCTGCGCCCATGTC 730
      ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 79398 TTAATTAATAGAAATGGCTCCCTTTGATAATTTCTCAACCCCTCTGCGCCCATGTC 79457
```

```
Query Match      60.1%; Score 602; DB 9; Length 4473;
Best Local Similarity 80.5%; Pred. No. 5.2e-176;
Matches 811; Conservative 0; Mismatches 170; Indels 27; Gaps 8;

Qy 1 GGAGTGGATTAACCGTGTAGTGGCCCTCAAGTGTGTGCGACCATGGAATGGGAGACTG 60
Db 2150 GGAGTGGACAAACCGTGTGGGTGCGCCCTCCAGGTGTGTGACCATGGAACGGAGACCG 2091

Qy 61 GAGGATACATGGATCCCAACTACAGCCCGCCAGCTCCTCCATGATGAGCCATGAGCCAGTT 120
Db 2090 GAGGATCCATGGATTAACCGGTGGGCTGTGACCTCCAGTACGAGCCATGAGCCAGCG 2031

Qy 121 GAATCTGAATGTGAAGTGAATGAAGACCGACGAGAGTACACTGAGCTCAACCTCAT 180
Db 2030 GAATCTGAATGCAAGACAGAAAGCGCCGACCGAGTCAATGACATCCAACCCCAT 1971

Qy 181 AACATGGGGTCA-GATCAAGAAACACACACGAGAGCTGAGAACTGTGTAGTGCCAGG 239
Db 1970 AACATGGGGACAAGATCAAGAAACCGACACAGAGAGCTGAGAACTACTGGAGCGCCAGG 1911

Qy 240 GTCAGGAAAAACCCCTGACTCCATGTTTATGGCCATGCTAGCTGTATATATCTGTGCG 299
Db 1910 GTCAGGAAAAACCCCTGACTCCATGTTTATGGCCATGCTAGCTGTATATATCTGTGCGG 1851

Qy 300 TATGATTTTCTGTGCGAAGCAAAACATATTTGGGCATATTTTCTTAACCCACCGGTAG 359
Db 1850 TATGTTTCCCTGTGTAGAGCAAAACATATTTGGGCATATTTTCTTAACCCACCTGGTAG 1791

Qy 360 TGTGA-TCATACTCTGAAGAGCACTCTCTCTGAGATATATCATGATCAAGAGCATCAG 418
Db 1790 TACGACTGTCTACTCTGGAGCAGACACTCTCTCTGAGATATATATGATCAGGAGCGTGGG 1731

Qy 419 TACCAGGACCTTAACTCCCTGACAGAGCAATAGACTCTCATACAAATGCTATCA 478
Db 1730 CACCAGGACCCCTAACTCCGCTGACACAGAG-----CAATAGCATCA 1688

Qy 479 ATTATACACTCCATTTGAGGAGACTCTCTTTATGTGTGTCACCCAGGATACATTTGCTCAACT 538
Db 1687 ATTACACGGCCCATTTGGAAGGACTCCCTTTATGTATCATCACCATGATACGTCACCTCACT 1628

Qy 539 CGAGTGGCTTGCAGTTGATCCCAAGCATGGTGTGAGTTACATCAATAAAATATGTATACC 598
Db 1627 GCAGCTGTCTTGTGATGCACTCAATCTCAGGCATGGTTAAAGTCACATGGAATAATATGTATT 1568

Qy 599 TATTAGACTTGTAGCTTATTAATATTAATCTGTGTAGTTACTAATCACTCTGCGCCCATC 658
Db 1567 TATTAGGCTCAGCTTATTAATGTACTGCGGTGTTCCACCAATCACTCTTAGGCCCCCATC 1508

Qy 659 -ACCCAAATTTACTGATT-ATACAGAAATGGGCTCCCTTTTGATTAATCTCA-CCGCCCTC 715
Db 1507 AACTCAAAATTTATGATTATATATGAATGGCTCCCTTTTGATTTCTACACCCCTC 1448

Qy 716 CTGGGCCCCATGCTCTGGCCCCCTTAGTACAAATAGTGCATGTTAATGGGAGACATTA 775
Db 1447 CTGGACCAAGTGTCTTGGCCCCCTTGGCCAGACAACTGCTATGTTAATGGGAAAC---A 1391

Qy 776 TTGACTGGGGTCCCTGTGGTCAATTAAGATGGAGAGATGAATCAGACCATGGCAT 835
Db 1390 TTGACTGGGGTCCCGTGGCCATTTAGAGGGGAGAGATGAATAATCAGACCTCTTGGGCACA 1331

Qy 836 AACTTCACTGGCACTGGTGGGAACTTTAAACATCTCTTCACTTCAACACACTGGGATTC 895
Db 1330 AACTTCACTGGCACTGGTGGGACCTTCAACACTCTTCACTACATCACTGGGATTC 1271

Qy 896 AAT--CCCAATCTGCCATGCAACTTGTGTGCGATGGAACGGCTTTAGCCACCTTTGCC 953
Db 1270 AGTCCCCAGTGTGGTCACAACTTGTGTGTGATAGGACAGGCTTTAAACCCGCCCTTACC 1211

Qy 954 TCAATGGCATTTATCAAGAAAGAGAGGTCCCAATTCAGGAGTCTATGTG 1001
Db 1210 TCAGTGACGTCTATAGAAACAGAGAGGCCCAATTCAGGAATTCGATATG 1163
```

RESULT 14  
AL451075/c  
LOCUS Human DNA sequence from clone RP11-215123 on chromosome 1, complete sequence.  
DEFINITION  
ACCESSION AL451075  
VERSION AL451075.15 GI:14148854  
KEYWORDS HTG.  
SOURCE human.  
ORGANISM Homo sapiens  
REFERENCE 1 (bases 1 to 153937)  
AUTHORS Coville,G.  
TITLE Direct Submission  
JOURNAL Submitted (26-MAY-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk  
requests: clonerequest@sanger.ac.uk  
COMMENT On May 18, 2001 this sequence version replaced gi:13751995.  
During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least one plasmid subclone or more than one M13 subclone; and the assembly was confirmed by restriction digest. The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: Em, EMBL; Sw, SWISSPROT; Tr, TREMBL; Wp, WORMPEP; Information on the WORMPEP database can be found at  
http://www.sanger.ac.uk/Projects/C\_elegans/wormpep This sequence was generated from part of bacterial clone contigs of human chromosome 1, constructed by the Sanger Centre Chromosome 1 Mapping Group. Further information can be found at  
http://www.sanger.ac.uk/HGP/Chrl  
RP11-215123 is from the library RPCI-11.1 constructed by the group of Pieter de Jong. For further details see  
http://www.chori.org/bacpac/home.htm  
VECTOR: PBAC3.6  
This sequence is the entire insert of clone RP11-215123 The true left end of clone RP11-545A16 is at 133805 in this sequence. The true right end of clone RP4-632K5 is at 55227 in this sequence.  
FEATURES  
Location/Qualifiers  
1..153937  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="1"  
/clone="RP11-215123"  
/clone.lib="RPCI-11.1"  
335..832  
/note="AluX repeat: matches 1..298 of consensus"  
1188..1484  
/note="AluJo repeat: matches 1..289 of consensus"  
1834..2550  
/note="L1MB8 repeat: matches 5346..6064 of consensus"  
2551..2851  
/note="AluSx repeat: matches 2..300 of consensus"  
2852..2891  
/note="L1MB8 repeat: matches 5307..5346 of consensus"  
2892..3198  
/note="AluSx repeat: matches 1..310 of consensus"  
3199..3957  
/note="L1MB8 repeat: matches 4474..5307 of consensus"  
4007..4177  
/note="AluSg/x repeat: matches 140..312 of consensus"  
4185..4488  
repeat\_region

```
repeat_region /note="AluSx repeat: matches 1. .304 of consensus"
repeat_region 4627. .4746
repeat_region /note="FLAM_C repeat: matches 7. .126 of consensus"
repeat_region 5157. .5466
repeat_region /note="AluJo repeat: matches 1. .309 of consensus"
repeat_region 5639. .5802
repeat_region /note="AluSg/x repeat: matches 164. .307 of consensus"
repeat_region 6305. .6344
repeat_region /note="20 copies 2 mer tt 85% conserved"
repeat_region 7423. .7510
repeat_region /note="44 copies 2 mer at 62% conserved"
repeat_region 8083. .8393
repeat_region /note="AluJ repeat: matches 1. .306 of consensus"
repeat_region 8431. .8734
repeat_region /note="AluY repeat: matches 1. .301 of consensus"
repeat_region 9829. .10142
repeat_region /note="AluSx repeat: matches 1. .312 of consensus"
misc_feature complement(10324. .10654)
repeat_region /note="match: STS: Em:G55659"
repeat_region 10810. .10977
repeat_region /note="MER53 repeat: matches 2. .189 of consensus"
repeat_region 11378. .11519
repeat_region /note="FLAM_C repeat: matches 1. .136 of consensus"
repeat_region 12045. .12346
repeat_region /note="AluJb repeat: matches 1. .299 of consensus"
repeat_region 12464. .12643
repeat_region /note="AluSp repeat: matches 125. .302 of consensus"
repeat_region 12729. .13015
repeat_region /note="AluSg repeat: matches 3. .298 of consensus"
repeat_region 13285. .13709
repeat_region /note="LTR repeat: matches 172. .691 of consensus"
repeat_region 13777. .14020
repeat_region /note="LTR repeat: matches 1. .256 of consensus"
repeat_region 14196. .14325
repeat_region /note="WLJ1 repeat: matches 3. .136 of consensus"
repeat_region 14739. .14870
repeat_region /note="AluSg/x repeat: matches 1. .132 of consensus"
repeat_region 15479. .15672
repeat_region /note="L2 repeat: matches 2130. .2325 of consensus"
repeat_region 16703. .16838
repeat_region /note="AluSg/x repeat: matches 1. .135 of consensus"
repeat_region 16842. .17132
repeat_region /note="AluSp repeat: matches 1. .299 of consensus"
repeat_region 18017. .18328
repeat_region /note="AluSp repeat: matches 1. .313 of consensus"
repeat_region 18370. .18546
repeat_region /note="AluJo repeat: matches 6. .189 of consensus"
repeat_region 18980. .19084
repeat_region /note="MIR repeat: matches 2. .109 of consensus"
repeat_region 19161. .19450
repeat_region /note="AluSg repeat: matches 2. .299 of consensus"
repeat_region 19464. .19777
repeat_region /note="AluSx repeat: matches 1. .307 of consensus"
repeat_region 19800. .19991
repeat_region /note="MIR repeat: matches 15. .227 of consensus"
repeat_region 19998. .20290
repeat_region /note="AluSg repeat: matches 1. .295 of consensus"
repeat_region 20959. .20994
repeat_region /note="AluSg repeat: matches 134. .169 of consensus"
repeat_region 21390. .21666
repeat_region /note="AluJb repeat: matches 1. .276 of consensus"
repeat_region 21671. .21714
repeat_region /note="Alu repeat: matches 247. .292 of consensus"
repeat_region 23092. .23166
repeat_region /note="FINGER2 repeat: matches 1. .75 of consensus"
repeat_region 23198. .23375
repeat_region /note="MER repeat: matches 1. .177 of consensus"
repeat_region 23496. .23772
repeat_region /note="AluJb repeat: matches 1. .291 of consensus"
repeat_region 24104. .24125
repeat_region /note="11 copies 2 mer at 100% conserved"
repeat_region 24820. .25115
repeat_region /note="AluSg repeat: matches 1. .296 of consensus"
```

```
repeat_region 25506. .25627
repeat_region /note="61 copies 2 mer ct 66% conserved"
repeat_region 26269. .26399
repeat_region /note="L2 repeat: matches 2152. .2294 of consensus"
repeat_region 26466. .26517
repeat_region /note="26 copies 2 mer tt 73% conserved"
repeat_region 26521. .26791
repeat_region /note="AluSg repeat: matches 1. .288 of consensus"
repeat_region 27124. .27416
repeat_region /note="AluSg repeat: matches 1. .293 of consensus"
repeat_region 27646. .27956
repeat_region /note="AluSx repeat: matches 1. .312 of consensus"
repeat_region 28612. .28841
repeat_region /note="L2 repeat: matches 1261. .1463 of consensus"
repeat_region 28842. .29132
repeat_region /note="AluJb repeat: matches 1. .304 of consensus"
repeat_region 29133. .29179
repeat_region /note="L2 repeat: matches 1463. .1513 of consensus"
repeat_region 29232. .29305
repeat_region /note="L2 repeat: matches 2598. .2671 of consensus"
repeat_region 29494. .29619
repeat_region /note="L2 repeat: matches 2617. .2750 of consensus"
repeat_region 29923. .30075
repeat_region /note="MER20 repeat: matches 68. .209 of consensus"
repeat_region 30076. .30200
repeat_region /note="FLAM_A repeat: matches 5. .128 of consensus"
repeat_region 30201. .30255
repeat_region /note="MER20 repeat: matches 12. .68 of consensus"
repeat_region 30360. .30658
repeat_region /note="AluSg repeat: matches 1. .298 of consensus"
repeat_region 32044. .32097
repeat_region /note="27 copies 2 mer tg 100% conserved"
repeat_region 32613. .32907
repeat_region /note="AluSp repeat: matches 3. .300 of consensus"
repeat_region 34314. .34560
repeat_region /note="AluSg repeat: matches 67. .313 of consensus"
repeat_region 35005. .35287
repeat_region /note="AluSg repeat: matches 1. .296 of consensus"
repeat_region 35865. .35914
repeat_region /note="MER53 repeat: matches 1. .50 of consensus"
repeat_region 35962. .36100
repeat_region /note="L2 repeat: matches 1734. .1873 of consensus"
repeat_region 36395. .36428
repeat_region /note="17 copies 2 mer at 88% conserved"
repeat_region 36483. .36510
repeat_region /note="14 copies 2 mer ta 92% conserved"
repeat_region 36925. .37886
repeat_region /note="match: STS: Em:G07212"
repeat_region 38823. .39115
repeat_region /note="AluSg repeat: matches 1. .301 of consensus"
repeat_region 39749. .39772
repeat_region /note="12 copies 2 mer tt 100% conserved"
repeat_region 39794. .39937
repeat_region /note="AluJb repeat: matches 1. .144 of consensus"
repeat_region 39980. .40265
repeat_region /note="AluJo repeat: matches 1. .297 of consensus"
repeat_region 40488. .40782
```

```
Query Match 59.0%; Score 590.2; DB 9; Length 153937;
Best Local Similarity 76.3%; Pred. No. 4.1e-172;
Matches 764; Conservative 0; Mismatches 233; Indels 4; Gaps 3;

Oy 1 GGAGATGATAAACGGTGTGAGTGCCTCAAGTTGTGTGCGACCATGAATGGGAGACTG 60
Db 123176 GGAGATGAGAAACCATGTGGTGCCTCAAGTGTGCGCATGAATGGGAGACTG 123117

Oy 61 GAGGATACATGGATCCCACTACAGGCCAGCTCCTCCAGTATGAGCATGAGCCAGTT 120
Db 123116 GAGGAACCCAGGGTGGCCCAACCATGGCCGGTCCCTCGGTACGAGCATGAGCCAGCT 123057

Oy 121 GAATCTGAATGTGAAGATGGAATGAAGACCGACGAGAGTGCACACTGACGTCAACCTCAT 180
Db 123056 GAGCCTGAGTGCAAAGACGGAGAGAGGCTGACAAAGGTCAGGTACATCAACCTCAT 122997
```

Qy 181 AACATGGGTGATCAAGAAACACACAGAGCTGAGAACTGGTGTAGTGCCAGG 240  
Db 122996 AACCTGGGACAACTCAAGAAACACATCAGAGAGCTGAGAACTGAGGATCAGG 122937  
Qy 241 TCAGCAAAACCCCTGACTCCATGTTTATGGCCATGCTGTAATATCTGTGGAGT 300  
Db 122936 ACAGCAAAACCCCTGACTCCATGTTTATGGCCATGCTGTAATATCTGTGGAGT 122877  
Qy 301 ATGATTTTCTGTGAGAGCAAAACATATTTGGGCATATTTCTTAACCCACCGGTAGT 360  
Db 122876 ATGTTTCTGTGAGAGCAAAACATATTTGGGCATATTTCTTAACCCACCGGTAGT 122817  
Qy 361 GTGATC-ATATCTGAAGCAGCTCCCTCCAGATATATATGATCAAGAGCATCAGT 419  
Db 122816 ACGACCTAACTTTGGAGTGAACCTCCCTCCAGATATATATGATCAAGAGCATCAGT 122757  
Qy 420 ACCAGGACCTTAATCCCTCCGACACAGAGCAATTAAGTCTCATACATATGTTATCA 479  
Db 122756 TCCAGGACCTTAATCCCTCCGACACAGAGCAATTAAGTCTCATACATATGTTATCA 122697  
Qy 480 TTATACCACTCCATGGAGGACTTCCCTTATGTTGTCACCCAGGATACATGCTCAACTG 539  
Db 122696 TTATACC-TTCCACTGGAGGACTTCCCTTATGTTGTCACCCAGGATACATGCTCAACTG 122638  
Qy 540 CAGTTGCCCTTCAGTTGATCCCAAGCATGGTTGAGTTTACCATAAATAATATGTAAT 599  
Db 122637 TAGCTGTCTTAACTTCAAGCTCAACATGTTGAGTCACTATGGAATAATATGTAAT 122578  
Qy 600 ATTAGACCTTAGCTTATATATATTAATTAATTAATTAATTAATTAATTAATTAAT 659  
Db 122577 ATTAGGCTTGTGTTCTATTAATTAATTAATTAATTAATTAATTAATTAATTAAT 122518  
Qy 660 CCCAAATGTACTGATTAACAGAGTGGCTCCCTTTCATATCTCACCCCTCCCTG 719  
Db 122517 CCCTAATGTGCTGATTAACAGATAGATTAATTAATTAATTAATTAATTAATTAAT 122458  
Qy 720 GGGCCACTGCTTGGCCCTTAGCTAGACAAATAGTCCATGTTAATGGAGACATTAATGA 779  
Db 122457 GGGCCAGTGTCTGGTCCACTGGCTAGAAAACAACTATGTTAATGGAGACATTAATGA 122398  
Qy 780 CTGGGTCCCTGTTGATTAAGATGGAGAGATGAGAAATCAGACCATGATGCAAACT 839  
Db 122397 TTGGAGACCTAAAGGCTAAATTAGATGGAAGAAATGAAATCAGAAATCAGACAACT 122338  
Qy 840 TCAGTGGACCTGGTGGGAAACTTTAATCTCTTCACTTCAACACACTGGGATTCATC 899  
Db 122337 TCGCTGGCATTTGATGAGAGCTTTAATGCTTCTTTTATACACCCGAGATCCAATC 122278  
Qy 900 CCAATCTGCCATGCAATTTGCTTGGCATGGAACGGCTTTAGCCCACTTTGCTTCAATG 959  
Db 122277 CCAGTCTGCCCGCAGATTTGCTTGGCATGGAACGGCTTTAGCCCACTTTTCTCTC--TG 122220  
Qy 960 GCATTATCAAGGAAGAGAGTCCCAATTCAGGAGTCTATGT 1000  
Db 122219 GCATTATCCAGGAGGAAGACCAATTCAGAGACGATAT 122179

RESULT 15  
AC008567/c  
LOCUS AC008567 153704 bp DNA linear PRI 06-DEC-2001  
DEFINITION Homo sapiens chromosome 19 clone CTC-543D15, complete sequence.  
ACCESSION AC008567  
VERSION AC008567.5 GI:17386227  
KEYWORDS HTC.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 153704)  
REFERENCE  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL

REFERENCE  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (03-AUG-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (22-APR-2000) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE  
AUTHORS DOE Joint Genome Institute and Stanford Human Genome Center.  
TITLE Direct Submission  
JOURNAL Submitted (06-DEC-2001) DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT On Dec 6, 2001 this sequence version replaced gi:7636348.  
Draft Sequence Produced by DOE Joint Genome Institute  
www.jgi.doe.gov  
Finishing Completed at Stanford Human Genome Center  
www.shgc.stanford.edu  
Quality: Phrap Quality >=40 99.8% of Sequence;  
Estimated Total Number of Errors is 0.3.  
STS Content:  
WI-12462 G13389.  
Location/Qualifiers  
1. 153704  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="19"  
/clone="CTC-543D15"  
BASE COUNT 45188 a 34926 c 31830 g 41760 t  
ORIGIN

Query Match 58.8%; Score 589; DB 9; Length 153704;  
Best Local Similarity 76.1%; Pred. No. 9.8e-172;  
Matches 765; Conservative 0; Mismatches 235; Indels 5; Gaps 3;

Qy 1 GGAGATGATAAACCGTGTGAGTGCCTCAAGTTGTGCGACCATGGAATGGGAGACTG 60  
Db 97990 GGAGATGACAAACCATTTGGTGTCTCAAGTGCATGCAACCATGGAACAGGAGACTG 97931  
Qy 61 GAGGATACATGGATGCCAATACAGGCCAGCTCTCCAGTATGAGCCATGAGCCAGTT 120  
Db 97930 GAGGATCCATGGCGGCCCAACCATGGCTCTGCTCCCTGATGGCCATGAGCCAGT 97871  
Qy 121 GAATCTGAATGTGAAGATGAAGATGAAGACCGAGAGTCACTGAGCTCAACCTCAT 180  
Db 97870 GAGCTGAGTGTGAAGATGGAGAAAGGCTGACCGAGTCAAGGAGCATCAACCCCAT 97811  
Qy 181 AACATGGGTGATCAAGAAACACACAGAGCTGAGAACTGGTGTAGTGCCAGG 240  
Db 97810 AACCTGGGACAACTCAAGAAACACACAGAGCTGAGAAATGCTGGAGCGTCAGG 97751  
Qy 241 TCAGCAAAACCCCTGACTCCATGTTTATGGCCATGCTGTAATATCTGTGGAGT 300  
Db 97750 CCAGCAAAACCCCTGACTCCATGTTTATGGCCATGCTGTAATATCTGTGGAGT 97691  
Qy 301 ATGATTTTCTGTGAGAGCAAAACATATTTGGGCATATTTCTTAACCCACCGGTAGT 360  
Db 97690 ATGTTTCTGTGAGAGCAAAACATATTTGGGCATATTTCTTAACCCCTAGTATG 97631  
Qy 361 GT-GATCATCTCTGAAGCAGCAGCTCCCTCCAGATATATATGATCAAGAGCATCAGT 419  
Db 97630 ATGGCTTGTACTTTGGAGTGGCAGCTCCCTCCAGATATATATGATCAAGAGCATGAGC 97571  
Qy 420 ACCAGGACCTTAATCCCTCCGACACAGAGCAATTAAGTCTCATACATATGTTATCA 479  
Db 97570 TCCAGGACCTTAATCCCTCCGACATGAACAGTATAGACTCTCAGAAATATGTCATCA 97511  
Qy 480 TTATACCACTCCATGGAGGAGCTTCCCTTATGTTGTCACCCAGGATACATGCTCAACTG 539  
Db 97510 TTATATGCCCCCTGAGGAGGACTCCCTTGTGTTATCACTACAAAGATATTTGCTCAACTG 97451





GenCore version 5.1.3  
Copyright (c) 1993 - 2002 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 16:24:42 ; Search time 1812.02 Seconds  
(without alignments)  
16077.036 Million cell updates/sec

Title: US-09-708-724a-3\_COPY\_99000\_100000  
Perfect score: 1001  
Sequence: 1 tggcagcgctgtagctccc.....ttttatccaccatcactaa 1001

Scoring table: IDENTITY\_NUC  
Gap 10.0, Gapext 1.0

Searched: 2054640 seqs, 14551402878 residues  
Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0  
Maximum DB seq length: 20000000000  
Post-processing: Minimum Match 0%  
Maximum Match 100%  
Listing first 45 summaries

Database : GenEmbl.\*

- 1: gb\_ba.\*
- 2: gb\_htg.\*
- 3: gb\_in.\*
- 4: gb\_em.\*
- 5: gb\_ov.\*
- 6: gb\_pat.\*
- 7: gb\_ph.\*
- 8: gb\_pl.\*
- 9: gb\_pr.\*
- 10: gb\_ro.\*
- 11: gb\_sts.\*
- 12: gb\_sy.\*
- 13: gb\_un.\*
- 14: gb\_vi.\*
- 15: em\_ba.\*
- 16: em\_fun.\*
- 17: em\_hum.\*
- 18: em\_in.\*
- 19: em\_mu.\*
- 20: em\_om.\*
- 21: em\_or.\*
- 22: em\_ov.\*
- 23: em\_pat.\*
- 24: em\_ph.\*
- 25: em\_pl.\*
- 26: em\_ro.\*
- 27: em\_sts.\*
- 28: em\_un.\*
- 29: em\_vi.\*
- 30: em\_htg\_hum.\*
- 31: em\_htg\_inv.\*
- 32: em\_htg\_other.\*
- 33: em\_htg\_mus.\*
- 34: em\_htg\_pln.\*
- 35: em\_htg\_rod.\*
- 36: em\_htg\_mam.\*
- 37: em\_htg\_vrt.\*
- 38: em\_sy.\*
- 39: em\_htgo\_hum.\*
- 40: em\_htgo\_mus.\*
- 41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID         | Description        |
|------------|-------|-------------|--------|----|------------|--------------------|
| c 1        | 1001  | 100.0       | 153629 | 9  | AC092325   | AC092325 Homo sapi |
| c 2        | 990   | 98.9        | 191768 | 9  | AC092357   | AC092357 Homo sapi |
| c 3        | 550.4 | 55.0        | 191014 | 9  | AC113268   | AC113268 Papio cyn |
| c 4        | 494.2 | 49.4        | 151049 | 9  | AC018558   | AC018558 Homo sapi |
| c 5        | 425.6 | 42.5        | 2719   | 9  | AK096292   | AK096292 Homo sapi |
| c 6        | 425.6 | 42.5        | 3702   | 9  | BC036456   | BC036456 Homo sapi |
| c 7        | 284.8 | 28.5        | 142978 | 9  | AC097717   | AC097717 Homo sapi |
| c 8        | 272.2 | 27.2        | 167133 | 2  | AC026186   | AC026186 Homo sapi |
| c 9        | 272.2 | 27.2        | 201061 | 9  | AC007353   | AC007353 Homo sapi |
| c 10       | 129.6 | 12.9        | 137496 | 9  | AC005288   | AC005288 Homo sapi |
| c 11       | 126   | 12.6        | 111187 | 2  | AC009283   | AC009283 Homo sapi |
| c 12       | 126   | 12.6        | 161494 | 2  | AC110495   | AC110495 Homo sapi |
| c 13       | 125   | 12.5        | 82453  | 9  | AC004558   | AC004558 Homo sapi |
| c 14       | 125   | 12.5        | 179040 | 2  | AC022732   | AC022732 Homo sapi |
| c 15       | 124.8 | 12.5        | 141594 | 9  | AC068302   | AC068302 Homo sapi |
| c 16       | 124.8 | 12.5        | 190588 | 2  | AC009861   | AC009861 Homo sapi |
| c 17       | 124.8 | 12.5        | 190610 | 2  | AC130419   | AC130419 Homo sapi |
| c 18       | 123.6 | 12.3        | 194995 | 9  | AC018696   | AC018696 Homo sapi |
| c 19       | 123.4 | 12.3        | 171523 | 9  | AC018452   | AC018452 Homo sapi |
| c 20       | 122.8 | 12.3        | 132290 | 9  | AL627402   | AL627402 Human DNA |
| c 21       | 122.8 | 12.3        | 144735 | 2  | AL161458   | AL161458 Homo sapi |
| c 22       | 122.2 | 12.2        | 146443 | 9  | AC007488   | AC007488 Homo sapi |
| c 23       | 122.2 | 12.2        | 163057 | 2  | AC012419   | AC012419 Homo sapi |
| c 24       | 122   | 12.2        | 39645  | 2  | AC110593   | AC110593 Homo sapi |
| c 25       | 122   | 12.2        | 166993 | 2  | AC127522   | AC127522 Homo sapi |
| c 26       | 122   | 12.2        | 194973 | 2  | AC024474   | AC024474 Homo sapi |
| c 27       | 121.8 | 12.2        | 175403 | 9  | AC093268   | AC093268 Homo sapi |
| c 28       | 121.8 | 12.2        | 178182 | 2  | AC114980   | AC114980 Homo sapi |
| c 29       | 121.8 | 12.2        | 178314 | 2  | AC019025   | AC019025 Homo sapi |
| c 30       | 121.6 | 12.1        | 213999 | 2  | AC051619   | AC051619 Homo sapi |
| c 31       | 121.4 | 12.1        | 110000 | 2  | AC021072_1 | Continuation (2 of |
| c 32       | 120.8 | 12.1        | 180490 | 2  | AC069152   | AC069152 Homo sapi |
| c 33       | 120.8 | 12.1        | 192254 | 2  | AC122133   | AC122133 Homo sapi |
| c 34       | 120.6 | 12.0        | 180194 | 9  | AC078850   | AC078850 Homo sapi |
| c 35       | 120.6 | 12.0        | 181531 | 9  | AC007690   | AC007690 Homo sapi |
| c 36       | 120.6 | 12.0        | 183451 | 9  | AC005972   | AC005972 Homo sapi |
| c 37       | 120.4 | 12.0        | 51407  | 9  | AL357556   | AL357556 Human DNA |
| c 38       | 120.4 | 12.0        | 86295  | 2  | AC078810   | AC078810 Homo sapi |
| c 39       | 120.4 | 12.0        | 100000 | 9  | AP000503S2 | AP000504 Homo sapi |
| c 40       | 120.4 | 12.0        | 110000 | 2  | AC125619_4 | Continuation (5 of |
| c 41       | 120.4 | 12.0        | 110000 | 2  | AC027294_2 | Continuation (3 of |
| c 42       | 120.4 | 12.0        | 124156 | 9  | AL596132   | AL596132 Human DNA |
| c 43       | 120.4 | 12.0        | 136493 | 9  | AL662899   | AL662899 Human DNA |
| c 44       | 120.4 | 12.0        | 161413 | 9  | AC119736   | AC119736 Homo sapi |
| c 45       | 120.4 | 12.0        | 163682 | 9  | HSDJ71117  | AL132713 Human DNA |

ALIGNMENTS

RESULT 1  
AC092325/c 153629 bp DNA linear PRI 02-NOV-2001  
LOCUS Homo sapiens chromosome 16 clone RP11-14K3, complete sequence.  
DEFINITION AC092325  
ACCESSION AC092325  
VERSION AC092325.2 GI:16596537  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 153629)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 16  
JOURNAL Unpublished

REFERENCE 2 (bases 1 to 153629)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA.

REFERENCE 3 (bases 1 to 153629)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (03-NOV-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On or before Nov 2, 2001 this sequence version replaced gi:7329393, gi:14589514.

COMMENT  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.  
-----  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.  
-----

FEATURES  
source  
1. 153629  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-14K3"  
BASE COUNT 45079 a 31598 c 31052 g 45900 t  
ORIGIN

Query Match 100.0%; Score 1001; DB 9; Length 153629;  
Best Local Similarity 100.0%; Pred. No. 4.5e-263;  
Matches 1001; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 TGGCAGCGCGCTGTAGTCCAGCTACTCAGGAGACTGAGCAGGAGAAATCGTGTGAACCC 60  
|||||  
Db 130091 TGGCAGCGCGCTGTAGTCCAGCTACTCAGGAGACTGAGCAGGAGAAATCGTGTGAACCC 130032  
|||||

QY 61 GGGAGACGGAGGTTGCAGTGAGCCAGATCGCGTCACTGCACCTCCAGCCTGGCGACAGAC 120  
|||||  
Db 130031 GGGAGACGGAGGTTGCAGTGAGCCAGATCGCGTCACTGCACCTCCAGCCTGGCGACAGAC 129972  
|||||

QY 121 GTTCCGTTTCAAAAGAAAAAATAATATTAATAAAGAAATAAATCCGGCGTGGCGG 180  
|||||  
Db 129971 GTTCCGTTTCAAAAGAAAAAATAATATTAATAAAGAAATAAATCCGGCGTGGCGG 129912  
|||||

QY 181 TGACATCAGTCTGTGCTTAATGCTCGCGCGGTACCGTCCGCGAGTCTCTTCT 240  
|||||  
Db 129911 TGACATCAGTCTGTGCTTAATGCTCGCGCGGTACCGTCCGCGAGTCTCTTCT 129852  
|||||

QY 241 GAGGACCCCGCCCGCCACATCTCGCGCTTCCAATAAGAGTTCAGGTTTTCGCGTCCGCGT 300  
|||||  
Db 129851 GAGGACCCCGCCCGCCACATCTCGCGCTTCCAATAAGAGTTCAGGTTTTCGCGTCCGCGT 129792  
|||||

QY 301 GGTGTGTTCTTCTGTGTCACAGGTTGGAACTGGAGATGCCCTCTTCTCTCTCAGGACA 360  
|||||  
Db 129791 GGTGTGTTCTTCTGTGTCACAGGTTGGAACTGGAGATGCCCTCTTCTCTCTCAGGACA 129732  
|||||

QY 361 GRACCATGAGCTAGCGGCGCGCGGTTCGGAAGCTCCCGCTCCGCCAAGGGCGCCT 420  
|||||  
Db 129731 GAACCATGAGCTAGCGGCGCGCGGTTCGGAAGCTCCCGCTCCGCCAAGGGCGCCT 129672  
|||||

QY 421 COTCAGAGCCGTCGTCGTCGCCGCCCTGCGGGAGCTGACCCGAGCGCAGTGCAGAGTTGC 480  
|||||  
Db 129671 COTCAGAGCCGTCGTCGTCGCCGCCCTGCGGGAGCTGACCCGAGCGCAGTGCAGAGTTGC 129612  
|||||

QY 481 TCCTCTTGGCGGTGACACAGGAGGAGCTGGAGCGCGGATCATCTCCAGGAAGCGGCTG 540  
|||||  
Db 129611 TCCTCTTGGCGGTGACACAGGAGGAGCTGGAGCGCGGATCATCTCCAGGAAGCGGCTG 129552  
|||||

QY 541 AGTAGGAATCGACGCGCCACATCTCTCTTTACCCGGGGATGTGCAGGATTACCGTGAA 600  
|||||  
Db 129551 AGTAGGAATCGACGCGCCACATCTCTCTTTACCCGGGGATGTGCAGGATTACCGTGAA 129492  
|||||

QY 601 ATCATGACTGCTATCTCTGCGAAATPACCAATGGGAAAAATGGAGTCTAGAAATATTGCC 660  
|||||  
Db 129491 ATCATGACTGCTATCTCTGCGAAATPACCAATGGGAAAAATGGAGTCTAGAAATATTGCC 129432  
|||||

QY 661 TCCATTTTAGCCACCGGTTCCCGAGTAGCTGATTGGGTGATGAAGTCTCCGGAACA 720  
|||||  
Db 129431 TCCATTTTAGCCACCGGTTCCCGAGTAGCTGATTGGGTGATGAAGTCTCCGGAACA 129372  
|||||

QY 721 AATGCGCTGCCATGATAGTTTCTGAAAAAGTAACATGTTTGGTTTCCCGAACAACAATAC 780  
|||||  
Db 129371 AATGCGCTGCCATGATAGTTTCTGAAAAAGTAACATGTTTGGTTTCCCGAACAACAATAC 129312  
|||||

QY 781 AGACTCTGGAGCTTTTAAGCAGCTTTATATGTTATAGTTAATGCTTTTAAGTCAGAGTA 840  
|||||  
Db 129311 AGACTCTGGAGCTTTTAAGCAGCTTTATATGTTAATGCTTTTAAGTCAGAGTA 129252  
|||||

QY 841 GTTTATCAAGAGAAATTTGAATGATGGAATAAGACTCCACAGACTCAATTTAGAT 900  
|||||  
Db 129251 GTTTATCAAGAGAAATTTGAATGATGGAATAAGACTCCACAGACTCAATTTAGAT 129192  
|||||

QY 901 GTCCAAATTTCTCATCTACTACAAATCATTTCCAGGAAGAAAAAGATAGGACCTTTTGAAA 960  
|||||  
Db 129191 GTCCAAATTTCTCATCTACTACAAATCATTTCCAGGAAGAAAAAGATAGGACCTTTTGAAA 129132  
|||||

QY 961 ATCTGATGATCGGCGCATGTGTTTTATCCACCACACTAA 1001  
|||||  
Db 129131 ATCTGATGATCGGCGCATGTGTTTTATCCACCACACTAA 129091  
|||||

RESULT 2  
AC092357 Homo sapiens chromosome 16 clone RP11-332P24, complete sequence.  
DEFINITION AC092357 AC023764  
ACCESSION AC092357.2 GI:15187270  
VERSION HTG.  
KEYWORDS Homo sapiens.  
SOURCE  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 191768)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 16  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 191768)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE 3 (bases 1 to 191768)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (15-AUG-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
On or before Aug 15, 2001 this sequence version replaced gi:7596817, gi:14589546.  
COMMENT  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than

1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.

## Sequence Quality Assessment:

This entry has been annotated with sequence quality  
estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than  
1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the  
GenBank flat file format but are available as part  
of this entry's ASN.1 file.

FEATURES  
    source  
        Location/Qualifiers  
            1..191768  
            /organism="Homo sapiens"  
            /db\_xref="taxon:9606"  
            /chromosome="16"  
            /clone="RP11-332P24"  
BASE COUNT   57149 a 39592 c 40150 g 54877 t  
ORIGIN

Query Match           98.9%; Score 990; DB 9; Length 191768;  
Best Local Similarity 99.9%; Pred. No. 5e-260;  
Matches 1001; Conservative 0; Mismatches 0; Indels 1; Gaps 1;  
  
Qy   1 TGGCAGCGCGCTGAGTCCAGCTACTCAGGAGACTGAGGAGGAGATCGCTTGAACCC 60  
      |||||  
Db 118193 TGGCAGCGCGCTGAGTCCAGCTACTCAGGAGACTGAGGAGGAGATCGCTTGAACCC 118242  
      |||||  
  
Qy   61 GGGAGACGAGGTTGCAGTGGCAAGATCGCTCACTGCACTCCAGCTCGCGGACAGAC 120  
      |||||  
Db 118243 GGGAGACGAGGTTGCAGTGGCAAGATCGCTCACTGCACTCCAGCTCGCGGACAGAC 118302  
      |||||  
  
Qy   121 GTTCCGTTTCAAAAGAAAAAATAATATTAATAAAAAAGATTAATAATCGCGCGT 180  
      |||||  
Db 118303 GTTCCGTTTCAAAAGAAAAAATAATATTAATAAAAAAGATTAATAATCGCGCGT 118362  
      |||||  
  
Qy   181 TGACATGACTCTGTGTTAATGCTCGCGCGCTACCTGCTCGCGAGTCTTCTTCT 240  
      |||||  
Db 118363 TGACATGACTCTGTGTTAATGCTCGCGCGCTACCTGCTCGCGAGTCTTCTTCT 118422  
      |||||  
  
Qy   241 GAGGA-CGCCCGCCCCACTCTCGCTTCCCAATAAGAGTTCAAGTTTTCGGTTCGCGG 299  
      |||||  
Db 118423 GAGGACCGCCCCCCCCACTCTCGCTTCCCAATAAGAGTTCAAGTTTTCGGTTCGCGG 118482  
      |||||  
  
Qy   300 TGGTTGCTGTTCTGCTGCCACAGTTGGAACTGGAGATGCTCTTCTCTCTCAGGAC 359  
      |||||  
Db 118483 TGGTTGCTGTTCTGCTGCCACAGTTGGAACTGGAGATGCTCTTCTCTCTCAGGAC 118542  
      |||||  
  
Qy   360 AGAACCATGAGCTAGCGGACGCGCGTTCGCAAGCTCCCGTCCGCCAACGGCGCC 419  
      |||||  
Db 118543 AGAACCATGAGCTAGCGGACGCGCGTTCGCAAGCTCCCGTCCGCCAACGGCGCC 118602  
      |||||  
  
Qy   420 TCCTCAGAGCGCTCGCTGCGCGCTGCGGAGCTACCGCGAGCGAGTTCAGAGTTG 479  
      |||||  
Db 118603 TCCTCAGAGCGCTCGCTGCGCGCTGCGGAGCTACCGCGAGCGAGTTCAGAGTTG 118662  
      |||||  
  
Qy   480 CTCCTGTTGGCGGTGACAGGAGGAGTGGAGCGCGGATCATCTCCAGGAACGGGCT 539  
      |||||  
Db 118663 CTCCTGTTGGCGGTGACAGGAGGAGTGGAGCGCGGATCATCTCCAGGAACGGGCT 118722  
      |||||  
  
Qy   540 GAGTAGGAACTGACCGCCACATCTCTTTACCGGGGATGTGAGGATACCGTGA 599  
      |||||  
Db 118723 GAGTAGGAACTGACCGCCACATCTCTTTACCGGGGATGTGAGGATACCGTGA 118782  
      |||||  
  
Qy   600 AATCATGACTCGTCTATCTGCGAATTAACCAATGGAAAAATGGAGTCTAGAAAAATTATTC 659  
      |||||  
Db 118783 AATCATGACTCGTCTATCTGCGAATTAACCAATGGAAAAATGGAGTCTAGAAAAATTATTC 118842  
      |||||  
  
Qy   660 CTCATTTTAGCCACCACCGGTTCCCCAGTAGCTGTATTGGGATGATGAAGTGTCTCCCGAAC 719

|||||  
Db 118843 CTCATTTTACCCACCGGTTCCCGAGTAGCTGTATTGGGTGATGAAGTCTCCGGAAC 118902  
Qy   720 AAATGCGCTGCCATGATAGTTTCTGAAAGATTAACATCTTTGGTTCCCAACACACATA 779  
Db 118903 AAATGCGCTGCCATGATAGTTTCTGAAAGATTAACATCTTTGGTTCCCAACACACATA 118962  
Qy   780 CAGACTCTGGAGCTTTTAAAGCACCTTTATATGTTTATAGTTTAAATGCTTTTAAAGTCAGAGT 839  
Db 118963 CAGACTCTGGAGCTTTTAAAGCACCTTTATATGTTTATAGTTTAAATGCTTTTAAAGTCAGAGT 119022  
Qy   840 AGTTTATCAAGAGAAATTTGAATGATTGGAATAAGGACTCCACAGCATCTAATTTGTAGA 899  
Db 119023 AGTTTATCAAGAGAAATTTGAATGATTGGAATAAGGACTCCACAGCATCTAATTTGTAGA 119082  
Qy   900 TGTCCAATCTCTCATACTACAATCATTTCCAGGAGGAGAAAGATAGGACCTTTGAAA 959  
Db 119083 TGTCCAATCTCTCATACTACAATCATTTCCAGGAGGAGAAAGATAGGACCTTTGAAA 119142  
Qy   960 AATCTGATGATCGGCCATGTGTTTTTATCCACCATCACTAA 1001  
Db 119143 AATCTGATGATCGGCCATGTGTTTTTATCCACCATCACTAA 119184

RESULT 3  
AC113268/c  
LOCUS           AC113268           191014 bp   DNA   linear   PRI 14-AUG-2002  
DEFINITION   Papio cynocephalus anubis clone rp41-22m16, complete sequence.  
ACCESSION   AC113268  
VERSION   AC113268.8   GI:22095124  
SOURCE   HTG.  
         olive baboon.  
ORGANISM   Papio cynocephalus anubis  
          Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
          Mammalia; Eutheria; Primates; Catarrhini; Cercopitheidae;  
          Cercopithecinae; Papio.  
REFERENCE   1 (bases 1 to 191014)  
AUTHORS   Zhou,L., Fu,Y., Wu,J., Eichler,E. and Roe,B.A.  
TITLE   Papio anubis BAC Clone rp41-22m16  
JOURNAL   Unpublished  
REFERENCE   2 (bases 1 to 191014)  
AUTHORS   Zhou,L., Fu,Y., Wu,J., Eichler,E. and Roe,B.A.  
TITLE   Direct Submission  
JOURNAL   Submitted (28-FEB-2002) Department Of Chemistry And Biochemistry,  
          The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
          OK 73019, USA  
REFERENCE   3 (bases 1 to 191014)  
AUTHORS   Zhou,L., Fu,Y., Wu,J., Eichler,E. and Roe,B.A.  
TITLE   Direct Submission  
JOURNAL   Submitted (06-AUG-2002) Department Of Chemistry And Biochemistry,  
          The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
          OK 73019, USA  
REFERENCE   4 (bases 1 to 191014)  
AUTHORS   Zhou,L., Fu,Y., Wu,J., Eichler,E. and Roe,B.A.  
TITLE   Direct Submission  
JOURNAL   Submitted (14-AUG-2002) Department Of Chemistry And Biochemistry,  
          The University Of Oklahoma, 620 Parrington Oval, Room 208, Norman,  
          OK 73019, USA  
COMMENT   On Aug 4, 2002 this sequence version replaced gi:21104942.  
----- Genome Center  
Center: Department Of Chemistry  
The University Of Oklahoma  
Center code:UOKNOR  
-----  
FEATURES  
    source  
        Location/Qualifiers  
            1..191014  
            /organism="Papio cynocephalus anubis"  
            /sub\_species="anubis"  
            /db\_xref="taxon:9555"  
            /clone="rp41-22m16"  
            /cclone\_lib="RPCI - 41 Male (Olive) Baboon BAC Library"  
BASE COUNT   54882 a 40134 c 39560 g 56438 t  
ORIGIN

```
Query Match      55.0%; Score 550.4; DB 9; Length 191014;
Best Local Similarity 82.6%; Pred. No. 1.9e-139;
Matches 699; Conservative 0; Mismatches 121; Indels 26; Gaps 5;

Qy 1 TGGCAGCGCGCTGTAGTCCCGAGCTACTCAGAGACTGAGGAGGAGAGATCCCTTGAACCC 60
Db 76991 TGGCACAAGCCCTGTAGTTCCAGCTAGTCCAGAGACTGAGGAGGAGAGATCCCTTGAACCC 76932

Qy 61 GGGAGCAGGAGGTTCAGTGAACCAAGATCGGTCTACCTGCACCTCAGCTGGC----GAC 116
Db 76931 GGGAGTGGAGGTTCAGTGAACCAAGATCGGCGCCTACCTACCTCAGCTGGCAGACAG 76872

Qy 117 AGACGTTCCGTTTCAAAAGAAAAAAT-----AATATTAAATAAAAGAAATAAAATCCGGC 171
Db 76871 AGACACTCCGTTTCAAAAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATA 76812

Qy 172 GCTGCGCGGTGACATCACTCTCTGCTGCTTAATGCTTCGCGCGGCTACCGTCTCGCGGAG 231
Db 76811 AATAAAATCCCGCGCTGCGCGGTGACGTTGATGTTTCAATGCGGCTTCCGTCAGTGCGAG 76752

Qy 232 TCTCTTTCTGAGGACCCCGCCCGACCTCCGCTTCCAATAAGGAGCTTCAGGTTTTCG 291
Db 76751 TCCCTTTCTGAGGAC-----CCCGACCTCTGCGCTTCCAATAAGGAGCTCAGGTTTTCG 76697

Qy 292 GGTCCCGCTGTGCTGTCTGCTGCGCACAGGTGGAAGTGGAGATGCCCTCTCTCTTCT 351
Db 76696 ACTCACTCGCGTGTGCTGTCTGCTGCGCACAGGTGGAATGGAGATGCCCTCTCTCTTCT 76637

Qy 352 CTCAGGACAGAACCATGAGCTAGCGGCGAGCGCGGTTCGGAAGCTCCCGCTCGCGCAA 411
Db 76636 CTCAGGACAGAACCATGAGCGGCGAGCGGAGAGCTGGCGCGGAAGCTCCCTCTCAGCCAC 76577

Qy 412 CGGGCGGCTCTCAGAGCGGTTCGCGCGCGGCTCGCGGGAGCTGACCGCGAGCGCAGTG 471
Db 76576 TGGGACCGCTCAGAGCCCGGCTGCTCGCGCTCGCGGGAGCTGACCGT-AGCGCTGTG 76518

Qy 472 CAGAGTTGCTCTGTGTGCGGTGACAGGAGGAGCTGGAGCGGCGGATCATCTCCAGGA 531
Db 76517 CAGAGTTGCTCTGTGTGCGGTGACAGGAGGAGCTGGAGTGGCGGATCATCTCCAGGA 76458

Qy 532 AGCGGCTGAGTAGGAATGAGCGCGGCACATCTCTCTTTACCGGGGATGTGCAGGAT 591
Db 76457 AGAGAGCTGAGTAGGAATGAGCGCGGCACATCTCTCTTTCCCGGGGATGTGCAGGAT 76398

Qy 592 TACCGTGAATATCATGACTGCTCATCTCGCAATTACCAATGGAAATTTGGAGTCTAGAA 651
Db 76397 TACTGTGAATATGACTGCTCATCTCGCAATTACCAATGGAAATTTGGAGTCTAGAA 76338

Qy 652 ATTATTGCTCCATTTTAGCCCCACCGTTCCCGAGTAGCTGTATTGGGGTGATGAAGTGC 711
Db 76337 ATTATTGAGCCATTTTAGCCCCACCGTTCCCGAATAGCAGTATTGGGTGATGAAGTGC 76278

Qy 712 TCC-----GGACAATGCGGTGCGCATGATGTTTTCTGAAAAGTAACATGTTT 760
Db 76277 TCCCAAGCACATTTTCACAAAATTCAGCTGCGCATGATGTTTTCTGAAAAGTAACATGTTT 76218

Qy 761 GGTTCCTCCAGAACACAATACAGACTCTGAGCTTTTAAGCACTTTATATGTTATTAGTT 820
Db 76217 GGTTCCTCTGAAGGCAATACAGATCTGAGCTTTTAAGCACTTTATATGTTATTAGTT 820

Qy 821 AATGCT 826
Db 76157 TGTGCT 76152

RESULT 4
AC018558
LOCUS AC018558 151049 bp DNA linear PRI 02-NOV-2001
DEFINITION Homo sapiens chromosome 16 clone RP11-80F22, complete sequence.
ACCESSION AC018558
VERSION AC018558.5 GI:16596530
KEYWORDS HTG.
```

```
SOURCE Homo sapiens.
ORGANISM Homo sapiens
REFERENCE 1 (bases 1 to 151049)
AUTHORS DOE Joint Genome Institute.
TITLE Sequencing of Human Chromosome 16
JOURNAL Unpublished
REFERENCE 2 (bases 1 to 151049)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (14-DEC-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
REFERENCE 3 (bases 1 to 151049)
AUTHORS DOE Joint Genome Institute.
TITLE Direct Submission
JOURNAL Submitted (02-NOV-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT On Nov 2, 2001 this sequence version replaced gi:9795566.
Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.
-----
Sequence Quality Assessment:
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.
All manually edited bases have been reduced to quality zero.
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.
Base-by-base quality values are not generally visible from the Genbank flat file format but are available as part of this entry's ASN.1 file.
-----
FEATURES             Location/Qualifiers
     source            1..151049
                        /organism="Homo sapiens"
                        /db_xref="taxon:9606"
                        /chromosome="16"
                        /clone="RP11-80F22"
BASE COUNT 46827 a 30783 c 30749 g 42690 t
ORIGIN
Query Match      49.4%; Score 494.2; DB 9; Length 151049;
Best Local Similarity 76.1%; Pred. No. 4.7e-124;
Matches 689; Conservative 0; Mismatches 178; Indels 38; Gaps 5;

Qy 132 AAAGAAAAAATAATATTAAATAAAGAAATAAATCCGCGCTCGCGGTCACATCAGTC 191
Db 40966 AATGGAATGAGATTAAACACACAAATAAATAATCCAGCGCTCGCGCTGACATCGCC 41025

Qy 192 TCTGCTGCTTAATGCTCGCGCGGCTACCGTCTCT-----GCGCA 230
Db 41026 TCTGAAGTGTGTCGCCGCACTGGCTTCTGGCTGGCGGCGGACAGTTGGCGCTC 41085

Qy 231 GTCTCTTTCTGAGGACACCCCGCCCGCACTCTCCGCGCTTCCAATAAGAGAGTTTCAGTTTTG 290
Db 41086 CTGCAGCAGTCCCAACCTCAGGACTCCCTACCTCCACCTCCAAACAGCTCGGCTCTTTG 41145

Qy 291 CGTCCGCGTGTGCTGCTCTCTGCTGCCACAGCTTGAAGTGGAGATGCTCTTCCTTC 350
Db 41146 CGGTCCGCGGCGCTGCTGCTCTGCTCTCAGCTTGAATTTGGAGATGCTCTGCTCTGC 41205

Qy 351 TCTCAGGACAGAACCATGAGCTTAGCGGACGCGGCTTCGCGAAGCTCCCGCTCCGCA 410
Db 41206 TCTCAGGACAGAACCATGAGAACACGACGCGGCGGTCGACGATGCTCTCCCAATCCGTCG 41265

Qy 411 ACGGCGCGCTCTCTCAGAGCC--GTCCGTGCCCGCGCTGCCGGAGCTGACCCGACGCGAG 469
```



Db 634 TTTTGTGAAAAGTAACATGTTTGGTGGCCCCAGAACACAACTACTGACTTTGGAGCTTTTAA 693

Qy 799 GCACCTTTATATGTTATAGTAAATGCTTTTA-----AGTCAGAGTAGTATTATCAAAAGGA 853

Db 694 GCACCTTATATGTTATAGTAAATGCTTTTAATTTAAGTCAGAACTGTTTATCAAAAGGA 753

Qy 854 AAATTTGAATGATTGGAATAAGGACTCCACAGCATCTAATTAATTTAGATGTCOAATCTCTT 913

Db 754 AAGTTTGAATGTTTGGAAATAAGGACTCCATAGCATCTAACTGTAGA--TCCAGCTCTTCT 811

Qy 914 CATACTACAATCATCTCCAGAGAGGAAAGATAGGACCTTTGAAAATCTGTATGGATCG 973

Db 812 CATACTAGCAATGTTGCGAGGAGAAAGTGGAGACCTGTGAAAATCTGTATGAGTCT 871

Qy 974 GCCATGCTGTTTTTATCCACCACACTAA 1001

Db 872 GCCATGATTTTTTATCCACCACACTAA 899

RESULT 6

BC036456 3702 bp mRNA linear PRI 12-AUG-2002

LOCUS Homo sapiens, clone MGC:33409 IMAGE:4825616, mRNA, complete cds.

DEFINITION BC036456

ACCESSION BC036456.1 GI:22209023

VERSION

KEYWORDS MGC.

SOURCE human.

ORGANISM Homo sapiens

Eukaryota; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

REFERENCE 1 (bases 1 to 3702)

AUTHORS Strausberg,R.

TITLE Direct Submission

JOURNAL Submitted (09-AUG-2002) National Institutes of Health, Mammalian Gene Collection (MGC), Cancer Genomics Office, National Cancer Institute, 31 Center Drive, Room 11A03, Bethesda, MD 20892-2590, USA

REMARK NIH-MGC Project URL: <http://mgc.nci.nih.gov>

COMMENT Contact: MGC help desk

Email: [cgaps-r@mail.nih.gov](mailto:cgaps-r@mail.nih.gov)

Tissue Procurement: Miklos Palkovits, M.D., Ph.D.

cDNA Library Preparation: Michael J. Brownstein (NHGRI) & Shiraki Toshiyuki and Piero Carninci (RIKEN)

cDNA Library Arrayed by: The I.M.A.G.E. Consortium (LLNL)

DNA Sequencing by: Institute for Systems Biology

<http://www.systemsbio.org>

contact: [amadansystemsbiology.org](mailto:amadansystemsbiology.org)

Anup Madan, Jessica Fahey, Erin Helton, Mark Kettelman, Anuradha Madan, Stephanie Rodrigues, Amy Sanchez and Michelle Whiting

Clone distribution: MGC clone distribution information can be found through the I.M.A.G.E. Consortium/LLNL at: <http://image.llnl.gov>

Series: IRAX Plate: 46 Row: n Column: 15

This clone was selected for full length sequencing because it passed the following selection criteria: GenomeScan gene prediction.

FEATURES

source

1. 3702

/organism="Homo sapiens"

/db\_xref="taxon:9606"

/clone="MGC:33409 IMAGE:4825616"

/tissue\_type="Testis"

/clone\_lib="NIH\_MGC\_97"

/lab\_host="DH10B"

/note="Vector: pBluescript"

487..1296

/codon\_start=1

/product="Unknown (protein for MGC:33409)"

/protein\_id="AAH36456.1"

/db\_xref="GI:22209024"

/translation="MTRHPENYQWENWSLENVATILAHFPNSYIWVVKSRMHLHFK SCYDNFVKSMFGAPEINTDFGAKHLYMLLVNFAFNLSQNSLKSUNWNKDSIASN CRSSPSHTTNGCQGEKVRTCEKSDSAMSFYPSLNDASFTLLGFSGKCVVLNQLLFE

CDS

LKEAKDKKNIDAFIKSIRTWYLDGGHSGSNWTWYTPVLKFEFAQTGIIIVHTVTPY QVDPNRKSWIGKEHKFVQILGDLGMQVTSQIHFTKEAPSIENHFRVHFYF"

BASE COUNT 1161 a 655 c 728 g 1158 t

ORIGIN

Query Match 42.5%; Score 425.6; DB 9; Length 3702;

Best Local Similarity 79.1%; Pred. No. 1.3e-105;

Matches 592; Conservative 0; Mismatches 124; Indels 32; Gaps 6;

Qy 284 GGTTTTGGGTGCGCGTGTGCTGCTCCGACAGGTTGGAACCTGGAGATGCTC 343

Db 153 GCTCCTCGGTGCGCGGTGCTGCTGCTCCGAGCTCGGAATCGGAACGCTC 212

Qy 344 TTCCTTCTCTCAGACAGAACCATG-AGCTAGCGGAGCGCGCTTCGGAGACTCCCC 402

Db 213 GTCTGTCTCTCAGCCAGAACCATGAACCGGGCGGAGCGCGCGCTCCT 272

Qy 403 CTCCGCCAACGGCGCTCTCAGAGCCGTCC--GTGCCGCCCTGCCGGGAGTGCACC 460

Db 273 CTGGCGGAGTGGCGCGCTCAGTGCCTGCAGCTTTCCACCGTGCCTGGAGCGGATCC 332

Qy 461 GCAGCGCAGTCCAGATGCTCTGTTGGCGGTGACAGGAGGAGGACTGGAGCGCGGAT 520

Db 333 GCAGCGCAGCAACGAATGCTCTGTTGGCGCGCGGAGGAGGACTGGAGCGCGAGGA 392

Qy 521 CATCTCCAGGAAGCGGCTGAGTAGGAAGTGCAGCCGC-----CACATCCTCTC 569

Db 393 CCTCCCGGGGACCCAGGAGGAGGAGCGCGCGCGCGCGCGCGCGCTCCTCTA 452

Qy 570 TTTACCGCGGAGTGCAGGANTACCGTGAATCATGCTGCTCATCTCGGAATACCA 629

Db 453 TTTCCCTGGGATGTCAGAAATTTACCATGAAATTTATGACTGCTCATCTGAGAATTTATCA 512

Qy 630 ATGGGAAATTTGGAGTCTAGAAATTTATGCTCCATTTAGCCACCCTGTCGCCAGTAG 689

Db 513 ATGGGAAATTTGGAGTCTAGAAATTTATGCTCCATTTTAGCCACCCTGTCGCCAGTAG 572

Qy 690 CTGTATTGGGTGATGAAGTCTCCGGAA-----CAAAATGGCTGCCATGATAG 738

Db 573 TTATATTGGGTGATAAATGTTCCCGAATGCAATTTGCACAAATTCAGCTGCTATGACAA 632

Qy 739 TTTTCTGAAAGTAACATGTTGTTGTTCCAGAGAACAAATACAGACTCTGGAGCTTTTAA 798

Db 633 TTTTGTGAAAGTAACATGTTGTTGTTCCCGCGCGCGCGCGCGCGCGCTTGGAGCTTTAA 692

Qy 799 GCACCTTTATATGTTATTAGTTAATGCTTTTA-----AGTCAGAGTAGTTTATCAAGGA 853

Db 693 GCACCTTTATATGTTATTAGTTAATGCTTTTAATTTAAGTCAGAAATAGTTTATCAAGGA 752

Qy 854 AAATTTGAATGATTGGAATAAGGACTCCACAGCATCTAATTTAGATGTCOAATCTCTTCT 913

Db 753 AAGTTTGAATGTTGGAATAAGGACTCCATAGCATCTAAGCTGTAGA--TCCAGTCTTCT 810

Qy 914 CATACTACAATCATCTCCAGAGAGGAAAGATAGGACCTTTGAAAATCTGTATGGATCG 973

Db 811 CATACTAGCAATGTTGCGAGGAGAAAGTGGAGACCTGTGAAAATCTGTATGAGTCT 870

Qy 974 GCCATGCTGTTTTTATCCACCACACTAA 1001

Db 871 GCCATGAGTTTTTATCCACCACACTAA 898

RESULT 7

AC097717 142978 bp DNA linear PRI 21-FEB-2002

LOCUS Homo sapiens BAC clone RP11-650N19 from 2, complete sequence.

DEFINITION AC097717 AC037452

ACCESSION AC097717.3 GI:18072229

VERSION

KEYWORDS HTG.

SOURCE Homo sapiens

ORGANISM Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

```

REFERENCE
AUTHORS      1 (bases 1 to 142978)
TITLE        Sulston, J.E. and Waterston, R.
JOURNAL      Toward a complete human genome sequence
MEDLINE      Genome Res. 8 (11), 1097-1108 (1998)
PUBMED       99063792
             9847074
REFERENCE
AUTHORS      2 (bases 1 to 142978)
TITLE        Tomlinson, C., Cotton, M. and Doeber, A.
JOURNAL      The sequence of Homo sapiens BAC clone RP11-650N19
REFERENCE
AUTHORS      3 (bases 1 to 142978)
TITLE        Waterston, R.H.
JOURNAL      Direct Submission
             Submitted (21-FEB-2001) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
REFERENCE
AUTHORS      4 (bases 1 to 142978)
TITLE        Waterston, R.H.
JOURNAL      Direct Submission
             Submitted (06-JAN-2002) Genome Sequencing Center, Washington
             University School of Medicine, 4444 Forest Park Parkway, St. Louis,
             MO 63108, USA
REFERENCE
AUTHORS      5 (bases 1 to 142978)
TITLE        Waterston, R.
JOURNAL      Direct Submission
             Submitted (21-FEB-2002) Department of Genetics, Washington
             University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA
             On Jan 6, 2002 this sequence version replaced gi:16756374.
COMMENT
-----
Center: Washington University Genome Sequencing Center
Center code: WUGSC
Web site: http://genome.wustl.edu/gsc
Contact: sapiens@watson.wustl.edu
-----
Center project name: H_NH0650N19
Drafting Center: WIBR
-----

NOTICE: This sequence may not represent the entire insert of this
clone. It may be shorter because we only sequence overlapping
clone sections once, or longer because we provide a small overlap
between neighboring data submissions.

This sequence was finished as follows unless otherwise noted:
all regions were double stranded, sequenced with an alternate
chemistry, or covered by high quality data (i.e., phred quality >=
30); an attempt was made to resolve all sequencing problems, such
as compressions and repeats; all regions were covered by sequence
from more than one subclone; and the assembly was confirmed by
restriction digest.

MAPPING INFORMATION:
Mapping information for this clone was provided by Dr. John D.
McPherson, Department of Genetics, Washington University, St. Louis
MO. For additional information about the map position of this
sequence, see http://genome.wustl.edu/gsc

SOURCE INFORMATION:
The RPCI-11 human BAC library was made from the blood of one male
donor, as described by Osoegawa, K., Woon, P.Y., Zhao, B., Frengen, E.,
Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved
approach for construction of bacterial artificial chromosome
libraries. Genomics 51:1-8. The clone may be obtained either from
Research Genetics, Inc. (http://www.resgen.com) or Pieter de Jong
and coworkers at http://www.chori.org
VECTOR: pBACE3.6

NEIGHBORING SEQUENCE INFORMATION:
The clone sequenced to the left is RP11-106I23; the clone sequenced
to the right is AC004464, 2000 bp overlap. Actual start of this
clone is at base position 1 of RP11-650N19.

Unsure base from position 41689 to 41691, and at position 41694.

```

Data from AC012529 was used to finish this clone, AC097717.  
Polymorphisms have been identified between AC097717 and AC004464.

The sequence of AC037452 has been incorporated into AC097717.

| FEATURES      | Location/Qualifiers   |
|---------------|---|
| source        | 1. 142978<br>/organism="Homo sapiens"<br>/db_xref="taxon:9606"<br>/chromosome="2"<br>/map="2"<br>/clone="RP11-650N19"<br>/clone_lib="RPCI-11" |
| repeat_region | 2. 574<br>/rpt_family="ERV1"  |
| repeat_region | 588. 689<br>/rpt_family="(TA)n"   |
| repeat_region | 689. 710<br>/rpt_family="(T)n"  |
| repeat_region | 996. 1036<br>/rpt_family="L1"   |
| repeat_region | 1037. 1332<br>/rpt_family="Alu"   |
| repeat_region | 1333. 1981<br>/rpt_family="L1"  |
| repeat_region | 2449. 2745<br>/rpt_family="Alu"   |
| repeat_region | 2857. 2923<br>/rpt_family="(TA)n"   |
| repeat_region | 3087. 3601<br>/rpt_family="MaLR"  |
| repeat_region | 4534. 4889<br>/rpt_family="MaLR"  |
| repeat_region | 5020. 5313<br>/rpt_family="L2"  |
| repeat_region | 5318. 5597<br>/rpt_family="L2"  |
| repeat_region | 5598. 5890<br>/rpt_family="Alu"   |
| repeat_region | 5891. 5962<br>/rpt_family="L2"  |
| repeat_region | 5963. 6262<br>/rpt_family="Alu"   |
| repeat_region | 6263. 7134<br>/rpt_family="L2"  |
| repeat_region | 7205. 7423<br>/rpt_family="L1"  |
| repeat_region | 7851. 8480<br>/rpt_family="L2"  |
| misc_feature  | 11144. 11147<br>/note="match to EST BE091529 (NID:g8481981)"  |
| repeat_region | 13428. 13448<br>/rpt_family="AT-rich"   |
| repeat_region | 13450. 13749<br>/rpt_family="Alu"   |
| repeat_region | 13956. 13984<br>/rpt_family="AT-rich"   |
| repeat_region | 13986. 14190<br>/rpt_family="L2"  |
| misc_feature  | 14971. 15380<br>/note="match to EST AA923721 (NID:g3068987) om39c01.s1"   |
| misc_feature  | 15055. 15374<br>/note="match to EST A1183644 (NID:g3734282) qf30f07.x1"   |
| repeat_region | 15849. 15921<br>/rpt_family="AT-rich"   |
| repeat_region | 16622. 16917<br>/rpt_family="Alu"   |
| misc_feature  | 17216. 17332<br>/note="match to EST A1183644 (NID:g3734282) qf30f07.x1"   |
| repeat_region | 17274. 17468<br>/rpt_family="ERV1"  |
| repeat_region | 17835. 17992  |



/rpt\_family="MIR"  
18015. .18057  
/rpt\_family="T-rich"  
18030. .18339  
/rpt\_family="Alu"  
18403. .18594  
/rpt\_family="MERI\_type?"  
19134. .19193  
/rpt\_family="Alu"  
19220. .19620  
/rpt\_family="L1"  
19622. .19729  
/rpt\_family="Alu"  
19732. .19770  
/rpt\_family="(GA)n"  
19760. .20656  
/rpt\_family="L1"  
20666. .20712  
/rpt\_family="Mariner"  
20739. .20799  
/rpt\_family="Alu"  
20812. .21084  
/rpt\_family="Alu"  
21121. .21377  
/rpt\_family="L1"  
21784. .22087  
/rpt\_family="7SLRNA"  
22118. .22425  
/rpt\_family="Alu"  
22481. .22639  
/rpt\_family="MaLR"  
22807. .22980  
/rpt\_family="MaLR"  
23092. .23546  
/rpt\_family="ERV1"  
23643. .24153  
/rpt\_family="MaLR"  
24183. .24352  
/rpt\_family="L2"  
24433. .24459  
/rpt\_family="(T)n"  
24825. .24946  
/rpt\_family="MaLR"  
24950. .25131  
/rpt\_family="L1"

Query Match 28.5%; Score 284.8; DB 9; Length 142978;  
Best Local Similarity 83.9%; Pred. No. 1.3e-66;  
Matches 364; Conservative 0; Mismatches 52; Indels 18; Gaps 3;  
Qy 584 TGCAGGATTACGCGTAAATCATGACTCGTCTCGCAATTCACCAATGGGAAATTTGGA 643  
Db 45768 TTCAGAATTACCATGAATTAATGACTCGTCTCGCAATTAATCAATGGGAAACTGGA 45827  
Qy 644 GTCTAGAAATATTGCTCCATTTTAGCCACCGGTTCCCGAGTACGTGTAATGGGGTGA 703  
Db 45828 GTCTAGAAATATTGCTCCATTTTAGCCACCGGTTCCCGAGTACGTGTAATGGGGTGA 45887  
Qy 704 TCAAGTGTCCGGAA-----CAAAATGCGTGCATGATAGTTTCTCGAAAGTGA 752  
Db 45888 TAAATGTCCGGAATGCAATTTGCGACAATTCAGCTGCTATGACAATTTGTGAAAAGTA 45947  
Qy 753 ACATGTTGGTTTCCCAACAACATACAGACTCTGGAGCTTTTAAAGCACCTTTATATGT 812  
Db 45948 ACATGTTGGTCCCAACAACATACACTGACTTTGGAGCTTTTAAAGCACCTTTATATGT 46007  
Qy 813 TATTAGTTAATGCT-----TTTAAGTCAGATAGTTTATCAAGGAAATTTGAATGATT 867  
Db 46008 TATTAGTTAATGCTTTTAAATTTAAGTCAGAATAGTTTATCAAGGAAATTTGAATGATT 46067  
Qy 868 GGAATAGGACGCCAGACATCTAATTTAGATGTCCAATTTCTCTCATACACAACTCA 927  
Db 46068 GGAATAGGACCTCCATAGCACTACTGTAGA--TCCAGTCTCTCTCATACCAATGG 46125

Qy 928 TTTCCAGGAAGAAAAGATAGACCTTTGAAAAATCTGATGATCGGCCATGTGTGTTTTA 987  
Db 46126 TTGCCAGGAGAAAAAGTGGAGACCTGTGAAAAATCTGATGAGTCTGCCATGAGTTTTTA 46185  
Qy 988 TCCACCATCACTAA 1001  
Db 46186 TCCACCATCACTAA 46199  
RESULT 8  
AC026186 167133 bp DNA linear HTG 29-MAY-2000  
Homo sapiens chromosome 3 clone RP11-303P8 map 3p, WORKING DRAFT  
SEQUENCE, 44 unordered pieces.  
AC026186  
AC026186.2 GI:8101183  
HTG; HTGS\_PHASE1; HTGS\_DRAFT.  
SOURCE  
Homo sapiens  
ORGANISM  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;  
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
1 (bases 1 to 167133)  
Baoy,J., Bao,Q., Bao,W., Bian,X., Cao,T., Chen,C., Chen,J., Ding,H.,  
Dong,W., Fan,H., Feng,X., Guan,Q., Gu,X., Guo,D., He,L., Hu,S.,  
Huang,F., Jin,Y., Kang,N., Li,C., Li,G., Li,J., Li,L.,  
Li,S., Li,T., Liu,Y., Liu,N., Liu,B., Liu,Y., Li,W., Li,W., Li,Y.,  
Luo,J., Niu,Y., Qi,Q., Qi,X., Song,S., Sun,M., Sun,W., Sun,Y.,  
Tao,R., Wang,H., Wang,J., Wang,L., Wang,L., Wang,R.,  
Wang,X., Wang,X., Wang,Y., Wu,D., Wu,Q., Xie,F., Xuan,Z., Xue,Y.,  
Yan,C., Yang,X., Yu,B., Zeng,Y., Zhang,G., Zhang,H., Zhang,H.,  
Zhang,L., Zhang,M., Zhang,X., Zhang,X., Zhang,Y., Zhang,Y.,  
Zhang,Z., Zhu,B., Yu,J. and Yang,H.  
Chromosome 3p genomic sequence  
Unpublished  
2 (bases 1 to 167133)  
Chen,J., Hu,S., Dong,W., Wang,J., Zhang,Y., Zhang,H., Liu,B.,  
Bao,W., Sun,Y., Wu,Q., Wang,H., Yang,X., Cheng,C., Wang,Y., Niu,Y.,  
Qi,X., Li,T., Zhang,H., Liu,N., Wu,D., Yu,B., Fan,H., Liu,Y.,  
Li,G., Li,C., Bao,Q., Bao,J., Wang,X., Song,L., Zhang,L., Guo,D.,  
Huang,F., Zhang,G., Li,J., Bian,X., Zhang,M., Li,L., Feng,X., Yu,J.  
and Yang,H.  
Direct Submission  
Submitted (21-MAR-2000) Human Genomic Center, Institute of  
Genetics, Chinese Academy of Sciences, Datun Road, Beijing, Beijing  
100101, P.R.China  
On May 29, 2000 this sequence version replaced gi:7272003.  
-----Genome Center  
Center:Beijing Center  
Center code:Beijing  
Website:http://hgsc.igtp.ac.cn  
http://www.genomics.org.cn  
Contact:hgsc@igtp.ac.cn  
----- Project Information  
Center project name:1# project  
Center clone name: RP11-303P8  
----- Summary Statistics  
Sequencing vector: pUC18; 100% of reads  
Chemistry: Dye-terminator; ET 55% of reads  
Assembly program: Phrap; version 0.990329  
Consensus quality: 144681 bases at least Q40  
Consensus quality: 152516 bases at least Q30  
Consensus quality: 159163 bases at least Q20  
Insert size: 121837; sum-of-contigs  
Quality coverage: 4.43x in Q20 bases,sum-of-contigs  
-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 44 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence





```
Db 96480 CACCACCCCTCCGCTCCCAATAGCTCGGGTTTTTTTTTTCGGGTGCTGCGGCTGCTGCT 96421
QY 311 CTTGTCGCCACAGTTTGAAGTGGAGATCCCTCTTCTCTCTCAGGACAGAACCATGAG 370
Db 96420 CTTACTCGCGCACTTGGAAATCCAAACGCTTCGTGGTCTCTCAGGACAGAACCATGAA 96361
QY 371 CTTAGCGGCGACCGCGCTTCGCGAAGCTCCCTCCGCAACAGGCGCTCTCTCAGAGCC 430
Db 96360 CCGAGC-GCAGCTGCGGTGCGTGACACCGCC-----AGGCACCGCTCGGAGGC 96312
QY 431 GTCCGTGCCCCGCTCCGCGGAGCTGACCCGAGCGCAGTGCAGAGTGTCTCTGTTGCC 490
Db 96311 TGACGTGTCTCTCTTGGCTGAGCCACCCGAGCGCAGCGGAGTTGCTCTCTGCTAGC 96252
QY 491 GGTGACACGAGGAGGACTGAGCGGGGATCATCTCCAGGAAGCGGCTGAGTAGGAAC 550
Db 96251 GGCTGCCGGGAGGACTGAGCGGGGACACCTTCGGGGAGCGTGCAAGGTGGAACC 96192
QY 551 CGAGCGCGCACATCTCTCTTTACCC-GGGGATGTCAGGATTACCGTGAAATCATGACT 609
Db 96191 CGAGCGCGCGCACCTCTCTCTCCCTGGGATGTGCAAAATTACCATGAATTTAGACT 96132
QY 610 CTTCTATCTGCAATACCAATGGGAAATTCGAGCTCAGAAATTTATGCTTCCATTTTA 669
Db 96131 TGTCTATCTGCGAGTTACAGTTGGAATTTGAGTCTAGAAATTTGTCGCCCATTTTA 96072
QY 670 GCCACCGGTTCCCAAGTACGTGTTGGGTGATGAAGTGTCTCGGAAAC-----719
Db 96071 GCCACCGGTTCCCAATGGCTATATTGGGTGATAAAGTGTCCGCAACGATTTCAC 96012
QY 720 -AAATGCGTGCATGATAGTTTCTGAAAGTACATGTTGGTTTCCGCAACACAT 778
Db 96011 AAATTCAGCTGCGGTGATAGTTTCTGAAAGTACATATTGGTTGCCGCAACACAT 95952
QY 779 ACAGACTCTGGAGCTTTTAAGCACCTTTATATGTTATTAGTTA 821
Db 95951 ACTACATATTAAATTTTATAACATTATATATATATATATA 95909

RESULT 9
AC007353/c
LOCUS Homo sapiens chromosome 16 clone RP11-488I20, complete sequence.
DEFINITION AC007353
ACCESSION AC007353.4 GI:14272747
VERSION HTG.
KEYWORDS Homo sapiens.
SOURCE Homo sapiens.
ORGANISM Homo sapiens.
REFERENCE 1 (bases 1 to 201061)
AUTHORS Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L., Bryant,J., Tesker,J., Meincke,L., Longmire,J., White,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R., McMurry,K., Han,C. and Deaven,L.
JOURNAL Direct Submission
TITLE Sequencing of Human Chromosome 16
AUTHORS Unpublished
2 (bases 1 to 201061)
AUTHORS Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D., Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L., Bryant,J., Tesker,J., Meincke,L., Longmire,J., White,S., Tatum,O., Campbell,C., Fawcett,J., Maltbie,M., Bussod,M., Sutherland,R., McMurry,K., Han,C. and Deaven,L.
JOURNAL Direct Submission
TITLE Submitted (22-APR-1999) Center for Human Genome Studies, DOE Joint Genome Institute, Los Alamos National Laboratory, MS M888, Los Alamos, NM 87545, USA
AUTHORS 3 (bases 1 to 201061)
JOURNAL Direct Submission
TITLE Submitted (01-JUN-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
AUTHORS On Jun 1, 2001 this sequence version replaced gi:9795555.
COMMENT -----Genome Center
Center: Joint Genome Institute
Center Code: JGI
```

Web site: <http://www.jgi.doe.gov>

#### Sequence Quality Assessment:

This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.

Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

#### FEATURES

Location/Qualifiers

1..201061  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-488I20"

BASE COUNT 57062 a 41994 c 41012 g 60993 t  
ORIGIN

Query Match 27.2%; Score 272.2; DB 9; Length 201061;  
Best Local Similarity 73.1%; Pred. No. 4e-63;  
Matches 426; Conservative 0; Mismatches 133; Indels 24; Gaps 5;

QY 252 CCCCACCTCTCCGCTTCCAATA-AGGAGTTTCAGGTTTTCGCGTGCCTGCTGCTGTT 310  
Db 45886 CACCACCTCCGCTCCCAATAGCTCGGGTTTTTTTTTTCGGTCTGTCGCGCTGCTGCT 45827

QY 311 CTTGTCGCCACAGTTTGAAGTGGAGATGCTTCCTCTCTCAGGACAGAACCATGAG 370  
Db 45826 CTTACTGCGCAACTTGAATCCAAACGCTTCGTGGTCTCTCAGGACAGAACCATGAA 45767

QY 371 CTTAGCGGCGGCTTCGCGAAGTCTCCCTCCGCAACGCGCTCTCTCAGAGCC 430  
Db 45766 CCGAGC-GCAGCTGCGGTGCGTGACACCGCC-----AGGCACCGCTTCGGAGGC 45718

QY 431 GTCCGTGCGCGCTTCGCGGAGCTGACCCGAGCGCATCTCCAGGACGCGGCTGAGTAGGAAC 550  
Db 45717 TGACGTGTCTCTCTTGGCTGAGCCACCCGAGCGCAGCGGCTGCTCTCTGCTAGC 45658

QY 491 GGTGACACGAGGAGGACTGAGCGGGGATCATCTCCAGGACGCGGCTGAGTAGGAAC 550  
Db 45657 GGCTCCGCGGAGGAGCTGAGCGCGGACACCTTCGCGGAGCGTGAAGGTGGAACC 45598

QY 551 CGAGCGCGCACATCTCTCTTTACCC-GGGGATGTCAGGATTACCGTGAATCATGACT 609  
Db 45597 CGAGCGCGCGCACCTCTCTCTCCCTGGGATGTGCAAAATTACCATGAATTTAGCT 45538

QY 610 CTTCTATCTGCAATACCAATGGGAAATTCGAGTCTAGAAATTTATGCTTCCCATTTTA 669  
Db 45537 TGTCTATCTCGAGTTACCAAGTTGGAATTTGAGTCTAGAAATTTGTCGCCCATTTTA 45478

QY 670 GCCACCGGTTCCCGAGTACGTGTTGGGTGATGAAGTGTCTCGGAAAC-----719  
Db 45477 GCCACCGGTTCCCGAATGGCTATATTGGGTGATAAAGTGTTCGCGAACGCAATTTCAC 45418

QY 720 -AAATGCGTGCATGATAGTTTCTGAAAGTAAACATGTTGGTTTCCGCAACACAT 778  
Db 45417 AAATTCAGCTGCGGTGATAGTTTCTGAAAGTAAACATATTGGTTGCCGCAACACAT 45358

QY 779 ACAGACTCTGGAGCTTTTAAGCACCTTTATATGTTATTAGTTA 821  
Db 45357 ACTACATATTAAATTTTATAACATTATATATATATATA 45315

#### RESULT 10

AC005288/c

LOCUS

DEFINITION Homo sapiens chromosome 17, clone hCIT.131\_K.11, complete sequence.

ACCESSION AC005288

VERSION AC005288.1 GI:3492893

KEYWORDS HTG.

137496 bp DNA linear PRI 29-AUG-1998

|           |   |  |                           |
|-----------|---|--|---------------------------|
| SOURCE    | Homo sapiens.   |  |                           |
| ORGANISM  | Homo sapiens  |  |                           |
| REFERENCE | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;   |  |                           |
| AUTHORS   | Mammalia; Eutheria; Primates; Catarrhini; Homnidae; Homo.   |  |                           |
| TITLE     | 1 (bases 1 to 137496)   |  |                           |
| JOURNAL   | Birren,B., Linton,L., Nusbaum,C. and Lander,E.  |  |                           |
| REFERENCE | Unpublished   |  |                           |
| AUTHORS   | 2 (bases 1 to 137496)   |  |                           |
|           | Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,   |  | /rpt_family="AT_rich"     |
|           | Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Boatn,C.,   |  | 9610. .9633               |
|           | Boutwell,C., Brown,A., Castle,A., Cerny,J., Cooke,P., Depayre,E.,   |  | /rpt_family="AT_rich"     |
|           | Devon,K., Dewar,K., Donelan,L., Etemadi,S., Ferreira,P.,  |  | 9640. .9939               |
|           | FitzHugh,W., Forrest,C., Funke,R., Gage,D., Gardyna,S.,   |  | /rpt_family="AluSg"       |
|           | Gensheimer,S., Geraghty,K., Gilmartin,T., Grant,G., Hagos,B.,   |  | complement(10233. .10312) |
|           | Harris,K., Horton,L., Howland,J.C., Hui,L., Jacotot,L., Kann,L.,  |  | /rpt_family="MIR"         |
|           | Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  |  | complement(10313. .10626) |
|           | Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,   |  | /rpt_family="AluSg"       |
|           | Nachman,A., Nahf,R., Naylor,J., Niloff,M., O'Connor,T., Pavlin,B.,  |  | complement(10627. .10779) |
|           | Peterson,K., Riley,R., Roberts,D., Rossello,R., Roy,A., Shyam,R.,   |  | /rpt_family="MIR"         |
|           | Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,  |  | complement(10869. .10942) |
|           | Strickland,C., Subramanian,A., Torruella-Miller,I., Vassiliev,H.,   |  | /rpt_family="AluJ/FLAM"   |
|           | Vo,A., Wagner,A., Wang,B., Wheeler,J., Wu,Y., Ye,W.J., Zhao,J. and  |  | 11407. .11546             |
|           | Zody,M.   |  | /rpt_family="AluSg/x"     |
| TITLE     | Direct Submission   |  | complement(11636. .11937) |
| JOURNAL   | Submitted (16-JUL-1998) Whitehead Institute/MIT Center for Genome   |  | /rpt_family="AluSg"       |
| REFERENCE | Research, 320 Charles Street, Cambridge, MA 02141, USA  |  | complement(11993. .12298) |
| AUTHORS   | 3 (bases 1 to 137496)   |  | /rpt_family="AluSg"       |
|           | Birren,B., Fasman,K., Linton,L., Nusbaum,C., Lander,E., Allen,N.,   |  | complement(13171. .13469) |
|           | Anderson,M., Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J.,  |  | /rpt_family="AluSx"       |
|           | Boutwell,C., Brown,A., Castle,A., Cerny,J., Colangelo,M.,   |  | 13754. .14055             |
|           | Collins,S., Collimore,A., Cooke,P., Corliss,D., Depayre,E.,   |  | /rpt_family="AluSx"       |
|           | Devon,K., Dewar,K., Donelan,L., Ferreira,P., FitzHugh,W.,   |  | 14058. .14333             |
|           | Forrest,C., Funke,R., Gage,D., Gardyna,S., Geraghty,K., Grant,G.,   |  | /rpt_family="AluSg"       |
|           | Hagos,B., Heaford,A., Herena,B., Horton,L., Howland,J.C.,   |  | 15277. .15305             |
|           | Jacotot,L., Jones,C., Kann,L., Karatas,A., Lenocky,J.,  |  | /rpt_family="AT_rich"     |
|           | Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,  |  | complement(15306. .15604) |
|           | Meldrim,J., Molla,M., Morris,W., Morrow,J., Mychaleckyj,J.,   |  | /rpt_family="AluSg"       |
|           | Nahf,R., Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P.,   |  | complement(15607. .15881) |
|           | Pavlin,B., Peterson,K., Riley,R., Roberts,D., Roy,A., Severy,P.,  |  | /rpt_family="AluSg"       |
|           | Stange-Thomann,N., Stilwell,J., Stojanovic,N., Stone,C.,  |  | complement(16147. .16454) |
|           | Subramanian,A., Tesfaye,S., Tichovolisky,N., Torruella-Miller,I.,   |  | /rpt_family="AluSg"       |
|           | Vassiliev,H., Vo,A., Wagner,A., Wheeler,J., Wu,Y., Wyman,D.,  |  | complement(16456. .16750) |
|           | Ye,W.J., Zhao,J. and Zody,M.  |  | /rpt_family="AluSg"       |
| TITLE     | Direct Submission   |  | 17102. .17164             |
| JOURNAL   | Submitted (29-AUG-1998) Whitehead Institute/MIT Center for Genome   |  | /rpt_family="MIR"         |
| REFERENCE | Research, 320 Charles Street, Cambridge, MA 02141, USA  |  | complement(17326. .17472) |
| COMMENT   | On Aug 29, 1998 this sequence version replaced gi:3449085.  |  | /rpt_family="FLAM_C"      |
|           | All repeats were identified using RepeatMasker: Smit, A.F.A. &  |  | complement(17561. .17845) |
|           | Green, P. (1996-1997)   |  | /rpt_family="AluSg1"      |
|           | <a href="http://ftp.genome.washington.edu/RM/RepeatMasker.html">http://ftp.genome.washington.edu/RM/RepeatMasker.html</a> . |  | complement(17877. .18033) |
| FEATURES  | Location/Qualifiers   |  | /rpt_family="AluSg"       |
| source    | 1. .137496  |  | complement(18034. .18337) |
|           | /organism="Homo sapiens"  |  | /rpt_family="AluSx"       |
|           | /db_xref="taxon:9606"   |  | complement(18338. .18456) |
|           | /chromosome="17"  |  | /rpt_family="AluSg"       |
|           | /map="17"   |  | complement(18770. .18794) |
|           | /clone="hCIT.131_K.11"  |  | /rpt_family="(TAAA)n"     |
|           | 1-1941"   |  | complement(18795. .19075) |
|           | /rpt_family="Research Genetics/Cal Tech CIB978SK-B (plates  |  | complement(19483. .19791) |
|           | 1-1941)"  |  | /rpt_family="AluSg"       |
|           | 1391. .1437   |  | 19832. .20136             |
|           | /rpt_family="purine-rich"   |  | /rpt_family="AluSc"       |
|           | complement(4525. .4815)   |  | complement(20142. .20437) |
|           | /rpt_family="AluSg"   |  | /rpt_family="AluSg"       |
|           | 5059. .5187   |  | complement(20670. .20734) |
|           | /rpt_family="AluJo/FLAM"  |  | /rpt_family="MER47A"      |
|           | 5189. .5442   |  | 20735. .21033             |
|           | /rpt_family="AluSx"   |  | /rpt_family="AluSg"       |
|           | 6296. .6354   |  | 21038. .21334             |
|           | /rpt_family="(CG)n"   |  | complement(21335. .21614) |
|           | 7179. .7209   |  | /rpt_family="MER47A"      |
|           | /rpt_family="gc-rich"   |  | 21961. .21982             |
|           | 9221. .9265   |  | /rpt_family="AT_rich"     |
|           |   |  | complement(21983. .22265) |
|           |   |  | /rpt_family="AluSp"       |
|           |   |  | complement(22281. .22303) |
|           |   |  | /rpt_family="(TAAA)n"     |
|           |   |  | complement(22304. .22548) |
|           |   |  | /rpt_family="AluSg1"      |

```
repeat_region complement(22567..22862)
repeat_region /rpt_family="AluSc"
repeat_region complement(23289..23350)
repeat_region /rpt_family="AluJ/FRAM"
repeat_region complement(23385..23541)
repeat_region /rpt_family="FRAM"
repeat_region complement(23636..23937)
repeat_region /rpt_family="AluSx"
repeat_region 24024..24067
repeat_region /rpt_family="AT_rich"
repeat_region complement(24068..24366)
repeat_region /rpt_family="AluSx"
repeat_region complement(24369..24660)
repeat_region /rpt_family="AluJo"
repeat_region 24661..24711
repeat_region /rpt_family="AT_rich"
repeat_region 24723..25021
repeat_region /rpt_family="AluSg"
repeat_region complement(25140..25272)
repeat_region /rpt_family="FLAM_A"
repeat_region complement(25427..25737)
repeat_region /rpt_family="AluSc"
repeat_region complement(25774..26059)
repeat_region /rpt_family="AluJo"
repeat_region complement(26150..26450)
repeat_region /rpt_family="AluJ"
repeat_region 26454..26487
repeat_region /rpt_family="AT_rich"
repeat_region complement(26498..26797)
repeat_region /rpt_family="AluSg"
repeat_region complement(26798..27092)
repeat_region /rpt_family="AluJb"
repeat_region complement(27113..27414)
repeat_region /rpt_family="AluJo"
repeat_region complement(27439..27734)
repeat_region /rpt_family="AluSg"
repeat_region complement(27812..28598)
repeat_region /rpt_family="L2"
repeat_region 28596
repeat_region /note="Probably A"
repeat_region complement(28623..28805)
repeat_region /rpt_family="AluJo"
repeat_region complement(28806..29111)
repeat_region /rpt_family="AluJo"
repeat_region complement(29112..29221)
repeat_region /rpt_family="AluJo"
repeat_region complement(29228..29364)
repeat_region /rpt_family="L1M4"
repeat_region complement(29375..29666)

Query Match 12.9% Score 129.6; DB 9; Length 137496;
Best Local Similarity 81.5% Pred. No. 4.8e-24;
Matches 150; Conservative 0; Mismatches 34; Indels 0; Gaps 0;

Oy 1 TGGCAGCGCCTGTAGTCCAGCTACTCAGGAGACTGAGGAGGAGAGACTCTTGAAACC 60
||||| |
Db 22717 TGGCGTGTCTGTAGTCCAGCTACTCAGGAGGCTGAGGAGGAGAGACTCTTGAAACC 22658
||||| |

Oy 61 GGGACAGGAGGTTCAGTGAGCCAGATCGGTCACCTGCACCTCCAGCTGGCGACAGAC 120
||||| |
Db 22657 GGGAGGAGAGGTTGACGTACGCCAAGATCGGCCACTGTCTCAGCTGACACAGAG 22598
||||| |

Oy 121 GTTCGGTTCAAAAGAAAAAATATATTAATAAAAGAAATAAAATCCGGCGCTCGCGGG 180
||||| |
Db 22597 ATTGCTCTCAAAAAAATAAAAAAACCACCACTAATCATGCTCTGCTGGCGCGG 22538
||||| |

Oy 181 TGAC 184
||| |
Db 22537 TGGC 22534

RESULT 11
AC009283/c
```

LOCUS  
DEFINITION

AC009283 111187 bp DNA linear HTG 19-AUG-2002  
Homo sapiens clone RP11-390P24, \*\*\* SEQUENCING IN PROGRESS \*\*\*, 2  
unordered pieces.

ACCESSION  
VERSION  
KEYWORDS  
SOURCE  
ORGANISM

AC009283 3 GI:22297433  
HTG; HTGS\_PHASE1; HTGS\_FULLTOP; HTGS\_ACTIVEFIN.  
human.  
Homo sapiens

REFERENCE  
AUTHORS  
TITLE  
JOURNAL  
REFERENCE  
AUTHORS

1 (bases 1 to 111187)  
Birren,B., Nusbaum,C. and Lander,E.  
Homo sapiens chromosome, clone RP11-390P24  
Unpublished  
2 (bases 1 to 111187)  
Birren,B., Linton,L., Barna,N., Nusbaum,C., Lander,E., Allen,N., Anderson,M.,  
Baker,J., Baldwin,J., Barna,N., Beckerly,R., Benn,J., Brown,A.,  
Castle,A., Cerny,J., Colangelo,M., Collins,S., Collymore,A.,  
Cooke,P., Dearellano,K., Depayre,E., Devon,K., Dewar,K.,  
Donelan,L., Doyle,M., Ferreira,P., FitzHugh,W., Forrest,C.,  
Funke,R., Gage,D., Galagan,J., Gardyna,S., Gilbert,D., Grant,G.,  
Hagos,B., Heaford,A., Horton,L., Howland,J.C., Jones,C., Kann,L.,  
Karatas,A., Lebecky,J., Lieu,C., Locke,K., Macdonald,P.,  
Marquis,N., McEwan,P., McGurk,A., McKernan,K., McLaughlin,J.,  
Meidrim,J., Mollam,M., Morris,W., Morrow,J., Mychaleckyj,J.,  
Naylor,J., Niloff,M., O'Connor,T., O'Donnell,P., Pavlin,B.,  
Peterson,K., Pollara,V., Riley,R., Roberts,D., Roy,A., Severy,P.,  
Stange-Thomann,N., Stojanovic,N., Stone,C., Subramanian,A.,  
Tesfaye,S., Toruella-Miller,I., Vassiliev,H., Vo,A., Wagner,A.,  
Wheeler,J., Wu,X., Wyman,D., Ye,W.J. and Zody,M.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Submitted (12-AUG-1999), Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
3 (bases 1 to 111187)  
Birren,B., Nusbaum,C., Lander,E., Ali,A., Allen,N., Anderson,S.,  
Barna,N., Bastien,V., Bloom,T., Boguslavsky,L., Boukhalter,B.,  
Camarata,J., Chang,J., Chazaro,B., Choepel,Y., Collymore,A.,  
Cook,A., Cooke,P., Dearellano,K., Dewar,K., Diaz,J.S., Dodge,S.,  
Farro,S., Ferreira,P., FitzGerald,M., Gage,D., Galagan,J.,  
Gardyna,S., Gord,S., Graham,L., Grand-Pierre,N., Hagos,B.,  
Horton,L., Hulme,W., Iliev,I., Johnson,R., Jones,C., Kamat,A.,  
Karatas,A., Kells,C., Landers,T., Levine,R., Lindblad-Toh,K.,  
Liu,G., Maclean,C., Macdonald,P., Major,J., Matthews,C.,  
McCarthy,M., Meldrim,J., Meneus,L., Mihova,T., Mlenga,V.,  
Murphy,T., Naylor,J., Nguyen,C., Nicol,R., Norbu,C., Norman,C.H.,  
O'Connor,T., O'Donnell,P., O'Neill,D., Oliver,J., Peterson,K.,  
Phunkhang,P., Pierre,N., Raymond,C., Retta,R., Rise,C., Rogov,P.,  
Roman,J., Roy,A., Schauer,S., Schuback,R., Seaman,S., Severy,P.,  
Smith,C., Spencer,B., Stange-Thomann,N., Stojanovic,N., Talamas,J.,  
Tesfaye,S., Theodore,J., Topham,K., Travers,M., Vassiliev,H.,  
Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Young,G., Zainoun,J.,  
Zembek,L., Zimmer,A. and Zody,M.

TITLE  
JOURNAL  
REFERENCE  
AUTHORS

Submitted (19-AUG-2002) Whitehead Institute/MIT Center for Genome  
Research, 320 Charles Street, Cambridge, MA 02141, USA  
On Aug 19, 2002 this sequence version replaced gi:17386471.  
All repeats were identified using RepeatMasker:  
Smit, A.F.A. & Green, P. (1996-1997)  
http://ftp.genome.washington.edu/RM/RepeatMasker.html

TITLE  
JOURNAL  
COMMENT

----- Genome Center  
Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR  
Web site: http://www-seq.wi.mit.edu  
Contact: sequence.submissions@genome.wi.mit.edu  
----- Project Information  
Center project name: L831  
Center clone name: 390\_P\_24  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 2 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.





```
misc_feature complement(17835..17938)
/feature="CGM6"
/notes="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: excellent, score: 75.000"
18196..18267
/rpt_family="(CA)n"
complement(20356..20573)
/rpt_family="AluSg"
complement(20574..20893)
/rpt_family="L1"
complement(20912..22752)
/rpt_family="L1BP1"
21269..21327
/rpt_family="(CA)n"
complement(23336..23488)
/rpt_family="MER54"
complement(23581..23869)
/rpt_family="AluSx"
complement(24094..24466)
/rpt_family="MLT1A2"
complement(24236..24326)
/notes="DPS similarity to (X16455) pCEA80-11 protein (647
AA) [Homo sapiens]."
pseudoc
complement(24472..24753)
/rpt_family="AluSx"
complement(25716..25842)
/rpt_family="L1MB4"
complement(25844..26146)
/rpt_family="AluSg"
complement(26147..26680)
/rpt_family="L1MB5"
complement(26722..27449)
/rpt_family="L1MB5"
27686..27900
/notes="predicted exon, program: grail2exons_human_1.3,
frame: 1, quality: good, score: 66.000"
28125..28216
/notes="predicted exon, program: grail2exons_human_1.3,
frame: 2, quality: excellent, score: 82.000"
complement(29034..29307)
/rpt_family="AluSg"
29574..29695
/rpt_family="MER20"
complement(30648..30860)
/rpt_family="L1"
complement(30935..31675)
/rpt_family="L1MB3"
31983..32149
/rpt_family="AluJb"
32551..32694
/rpt_family="MER4D"
32749..34381
/rpt_family="SVA"
34593..34758
/rpt_family="MER57_internal"
34825..35547
/rpt_family="MER4D"
35663..35731
/rpt_family="MER57_internal"
36428..36624
/rpt_family="MER57_internal"
36633..37326

Query Match 12.58; Score 125; DB 9; Length 82453;
Best local Similarity 84.8; Pred. No. 7.7e-23;
Matches 140; Conservative 0; Mismatches 25; Indels 0; Gaps 0;

QY 1 TGGCAGCGGCTGTAGTCCAGCTACTCAGGAGCTGAGGAGGAGTCCGCTTGAACCC 60
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 15538 TGGCATGCACTGTAGTCCAGCTACTCAGGAGGCTGAGGAGGAGTCCGCTTGAACCC 15597
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
QY 61 GGGAGACGGAGGTTGCAGTGCAGCCAAAGATCGCTGCTACTGCTCCAGCTCCAGCCTGGCGACAGAC 120
```

```
Db 15598 GGGAGGACAGGTTCCAGTGCAGCTGAGTGCAGCTGCAGCTCCAGCTCCAGCCTGGCCACAGAG 15657
QY 121 GTTCGGTTTCAAAAGAAAAAATAATATTAATAAAGAAATAAAAA 165
||||| ||||||| ||||||| ||||||| ||||||| ||||||| ||||||| |||||||
Db 15658 ACTCGTCTCAAAAAAAGAAAAAAGAAAAAAGAAAAAAGAAAAAGA 15702

RESULT 14
AC022732 179040 bp DNA linear HTG 24-AUG-2002
LOCUS Homo sapiens chromosome 17 clone RP11-399018 map 17, WORKING DRAFT
DEFINITION SEQUENCE, 32 unordered pieces.
AC022732
VERSION AC022732.3 GI:8347988
KEYWORDS HTG; HTGS_PHASE1; HTGS_DRAFT.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 179040)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Boquslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G., Castle,A.,
Choepel,X., Colangelo,M., Collins,S., Collamore,A., Cooke,P.,
DeArellano,K., Dewar,K., Domino,M., Doyle,M., Fenesor,J.,
Ferreira,P., FitzHugh,W., Forrest,C., Gage,D., Galagan,J.,
Gardyna,S., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Johnson,R., Jones,C., Kann,L., Karatas,A., Klein,J.,
Landers,T., Lehoczy,J., Levine,R., Lieu,C., Liu,G., Locke,K.,
Macdonald,P., Marquis,N., McEwan,P., McGurk,A., McKernan,K.,
McPheeters,R., Meldrim,J., Meneus,L., Morrow,J., Naylor,J.,
Norman,C.H., O'Connor,T., O'Donnell,P., Oliver,T.M., Peterson,K.,
Pierre,N., Pisani,C., Pollara,V., Raymond,C., Riley,R., Rothman,D.,
Roy,A., Santos,R., Severy,P., Spencer,B., Stange-Thomann,N.,
Stojanovic,N., Subramanian,A., Talamas,J., Tesfaye,S., Theodore,J.,
Tirrell,A., Vassiliev,H., Viel,R., Vo,A., Wu,X., Wyman,D., Ye,W.J.,
Zimmer,A. and Zody,M.
Direct Submission
Submitted (06-FEB-2000) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
3 (bases 1 to 179040)
Birren,B., Linton,L., Nusbaum,C., Lander,E., Abraham,H., Allen,N.,
Anderson,S., Baldwin,J., Barna,N., Bastien,V., Bida,F.,
Boguslavkiy,L., Boukhgalter,B., Brown,A., Burkett,G.,
Campopiano,A., Castle,A., Choepel,X., Colangelo,M., Collins,S.,
Collamore,A., Cooke,P., DeArellano,K., Dewar,K., Diaz,J.S.,
Dodge,S., Domino,M., Doyle,M., Ferreira,P., FitzHugh,W., Gage,D.,
Galagan,J., Gardyna,S., Ginde,S., Goyette,M., Graham,L.,
Grand-Pierre,N., Grant,G., Hagos,B., Heaford,A., Horton,L.,
Howland,J.C., Iliev,I., Johnson,R., Jones,C., Kann,L., Karatas,A.,
Klein,J., LaRocque,K., Lamazares,R., Landers,T., Lehoczy,J.,
Levine,R., Lieu,C., Liu,G., Locke,K., Macdonald,P., Marquis,N.,
McCarthy,M., McEwan,P., McGurk,A., McKernan,K., McPheeters,R.,
Meldrim,J., Meneus,L., Mihova,T., Miranda,C., Mlenga,V., Morrow,J.,
Murphy,T., Naylor,J., Norman,C.H., O'Connor,T., O'Donnell,P.,
O'Neil,D., Oliver,T.M., Oliver,J., Peterson,K., Pierre,N.,
Pisani,C., Pollara,V., Raymond,C., Riley,R., Rogov,P., Rothman,D.,
Roy,A., Santos,R., Schauer,S., Severy,P., Spencer,B.,
Stange-Thomann,N., Stojanovic,N., Subramanian,A., Talamas,J.,
Tesfaye,S., Theodore,J., Tirrell,A., Travers,M., Triggillo,J.,
Vassiliev,H., Viel,R., Vo,A., Wilson,B., Wu,X., Wyman,D., Ye,W.J.,
Young,G., Zainoun,J., Zimmer,A. and Zody,M.
Direct Submission
Submitted (24-AUG-2002) Whitehead Institute/MIT Center for Genome
Research, 320 Charles Street, Cambridge, MA 02141, USA
On Jun 8, 2000 this sequence version replaced gi:7767805.
All repeats were identified using RepeatMasker:
Smit, A.F.A. & Green, P. (1996-1997)
```

http://ftp.genome.washington.edu/RM/RepeatMasker.html  
----- Genome Center

Center: Whitehead Institute/ MIT Center for Genome Research  
Center code: WIBR

Web site: http://www-seq.wi.mit.edu

Contact: sequence.submissions@genome.wi.mit.edu

----- Project Information

Center project name: L5546

Center clone name: 399.O.18

----- Summary Statistics

Sequencing vector: M13; M7815; 100% of reads

Chemistry: Dye-terminator Big Dye; 100% of reads

Assembly program: Phrap; version 0.960731

Consensus quality: 164115 bases at least Q40

Consensus quality: 170430 bases at least Q30

Consensus quality: 173222 bases at least Q20

Insert size: 188000; agarose-fp

Insert size: 175940; sum-of-contigs

Quality coverage: 5.2 in Q20 bases; agarose-fp

Quality coverage: 5.6 in Q20 bases; sum-of-contigs

-----  
\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 32 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.  
-----

1 1289: contig of 1289 bp in length  
\* 1290 1389: gap of 100 bp  
\* 1390 2987: contig of 1598 bp in length  
\* 2988 3087: gap of 100 bp  
\* 3088 4215: contig of 1128 bp in length  
\* 4216 4315: gap of 100 bp  
\* 4316 5772: contig of 1457 bp in length  
\* 5773 5872: gap of 100 bp  
\* 5873 8065: contig of 2193 bp in length  
\* 8066 8165: gap of 100 bp  
\* 8166 10389: contig of 2224 bp in length  
\* 10390 10489: gap of 100 bp  
\* 10490 12485: contig of 1996 bp in length  
\* 12486 12585: gap of 100 bp  
\* 12586 15882: contig of 3297 bp in length  
\* 15883 15982: gap of 100 bp  
\* 15983 19042: contig of 3060 bp in length  
\* 19043 19142: gap of 100 bp  
\* 19143 21822: contig of 2680 bp in length  
\* 21823 21922: gap of 100 bp  
\* 21923 24967: contig of 3045 bp in length  
\* 24968 25067: gap of 100 bp  
\* 25068 29273: contig of 4206 bp in length  
\* 29274 29373: gap of 100 bp  
\* 29374 32046: contig of 2673 bp in length  
\* 32047 32146: gap of 100 bp  
\* 32147 35253: contig of 3107 bp in length  
\* 35254 35353: gap of 100 bp  
\* 35354 39343: contig of 3990 bp in length  
\* 39344 39443: gap of 100 bp  
\* 39444 43061: contig of 3618 bp in length  
\* 43062 43161: gap of 100 bp  
\* 43162 46849: contig of 3688 bp in length  
\* 46850 46949: gap of 100 bp  
\* 46950 48108: contig of 1159 bp in length  
\* 48109 48208: gap of 100 bp  
\* 48209 52614: contig of 4406 bp in length  
\* 52615 52714: gap of 100 bp  
\* 52715 58023: contig of 5309 bp in length  
\* 58024 58123: gap of 100 bp  
\* 58124 62500: contig of 4777 bp in length  
\* 62501 63000: gap of 100 bp  
\* 63001 67104: contig of 4104 bp in length  
\* 67105 67204: gap of 100 bp

\* 67205 73204: contig of 6000 bp in length  
\* 73205 73304: gap of 100 bp  
\* 73305 79910: contig of 6606 bp in length  
\* 79911 80010: gap of 100 bp  
\* 80011 83819: contig of 3809 bp in length  
\* 83820 83919: gap of 100 bp  
\* 83920 92133: contig of 8214 bp in length  
\* 92134 92233: gap of 100 bp  
\* 92234 101983: contig of 9750 bp in length  
\* 101984 102083: gap of 100 bp  
\* 102084 114101: contig of 12018 bp in length  
\* 114102 114201: gap of 100 bp  
\* 114202 126171: contig of 11970 bp in length  
\* 126172 126271: gap of 100 bp  
\* 126272 140708: contig of 14437 bp in length  
\* 140709 140808: gap of 100 bp  
\* 140809 160145: contig of 19337 bp in length  
\* 160146 160245: gap of 100 bp  
\* 160246 179040: contig of 18795 bp in length.  
FEATURES  
Location/Qualifiers  
1..179040  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="17"  
/map="17"  
/clone="RP11-399O18"  
/clone\_lib="RPC1-11 Human Male BAC"  
1..1289  
/note="assembly\_fragment"  
misc\_feature  
1390..2987  
/note="assembly\_fragment"  
misc\_feature  
3088..4215  
/note="assembly\_fragment"  
misc\_feature  
4316..5772  
/note="assembly\_fragment"  
misc\_feature  
5873..8065  
/note="assembly\_fragment"  
misc\_feature  
8166..10389  
/note="assembly\_fragment"  
misc\_feature  
10490..12485  
/note="assembly\_fragment"  
misc\_feature  
12586..15882  
/note="assembly\_fragment"  
misc\_feature  
15983..19042  
/note="assembly\_fragment"  
misc\_feature  
19143..21822  
/note="assembly\_fragment"  
misc\_feature  
21923..24967  
/note="assembly\_fragment"  
misc\_feature  
25068..29273  
/note="assembly\_fragment"  
misc\_feature  
29374..32046  
/note="assembly\_fragment"  
misc\_feature  
32147..35253  
/note="assembly\_fragment"  
misc\_feature  
35354..39343  
/note="assembly\_fragment"  
misc\_feature  
39444..43061  
/note="assembly\_fragment"  
misc\_feature  
43162..46849  
/note="assembly\_fragment"  
misc\_feature  
46950..48108  
/note="assembly\_fragment"  
clone\_end:SP6

Query Match 12.5%; Score 125; DB 2; Length 179040;  
Best Local Similarity 80.7%; Pred. No. 9.3e-23;  
Matches 146; Conservative 0; Mismatches 35; Indels 0; Gaps 0;

OY 1 TGGCACGCGCTGTAGTCCACGACTACTCAGGAGACTGAGGAGGAGGAGATCGCTTGAACCC 60  
|||||  
Db 150157 TGGCACGCGCTGTAGTCCACGACTACTCAGGAGCTGAGGAGGAGGATCGCTTCAACT 150216  
|||||







GenCore version 5.1.3  
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: December 24, 2002, 16:24:42 ; Search time 1812.02 Seconds  
(without alignments)  
16077.036 Million cell updates/sec

Title: US-09-708-724A-3\_COPY\_50000\_51000

Perfect score: 1001

Sequence: 1 agcaactgttaagtctgggc.....ggccctgctgcatgtgacc 1001

Scoring table: IDENTITY\_NUC

Gapop 10.0 , Gapext 1.0

Searched: 2054640 seqs, 1451402878 residues

Total number of hits satisfying chosen parameters: 4109280

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

GenEmbl.\*

1: gb\_ba.\*

2: gb\_hgt.\*

3: gb\_in.\*

4: gb\_om.\*

5: gb\_ov.\*

6: gb\_pat.\*

7: gb\_ph.\*

8: gb\_pl.\*

9: gb\_pr.\*

10: gb\_ro.\*

11: gb\_sts.\*

12: gb\_sy.\*

13: gb\_un.\*

14: gb\_vi.\*

15: em\_ba.\*

16: em\_fun.\*

17: em\_hum.\*

18: em\_in.\*

19: em\_mu.\*

20: em\_om.\*

21: em\_or.\*

22: em\_lo.\*

23: em\_pat.\*

24: em\_ph.\*

25: em\_pl.\*

26: em\_ro.\*

27: em\_sts.\*

28: em\_un.\*

29: em\_vi.\*

30: em\_htg\_hum.\*

31: em\_htg\_inv.\*

32: em\_htg\_other.\*

33: em\_htg\_mus.\*

34: em\_htg\_pln.\*

35: em\_htg\_rod.\*

36: em\_htg\_man.\*

37: em\_htg\_vrt.\*

38: em\_sy.\*

39: em\_htgo\_hum.\*

40: em\_htgo\_mus.\*

41: em\_htgo\_other.\*

Pred. No. is the number of results predicted by chance to have a

score greater than or equal to the score of the result being printed,  
and is derived by analysis of the total score distribution.

SUMMARIES

| Result No. | Score | Query Match | Length | DB | ID       | Description |
|------------|-------|-------------|--------|----|----------|-------------|
| 1          | 1001  | 100.0       | 151049 | 9  | AC018558 | Homo sapi   |
| 2          | 1001  | 100.0       | 191768 | 9  | AC092357 | Homo sapi   |
| C 3        | 659.2 | 65.9        | 131014 | 9  | AC113268 | Papio cyn   |
| C 4        | 646.2 | 64.6        | 214025 | 9  | AC007882 | Homo sapi   |
| C 5        | 646.2 | 64.6        | 227856 | 2  | AC007908 | Homo sapi   |
| C 6        | 644.6 | 64.4        | 70048  | 9  | HSU91320 | Human chrom |
| C 7        | 641.4 | 64.1        | 129290 | 2  | AC020761 | Homo sapi   |
| C 8        | 641.4 | 64.1        | 167553 | 2  | AC114938 | Homo sapi   |
| C 9        | 641.4 | 64.1        | 185994 | 2  | AC002042 | Homo sapi   |
| C 10       | 589.4 | 58.9        | 201061 | 9  | AC007353 | Homo sapi   |
| C 11       | 569.6 | 56.9        | 144577 | 9  | AC023824 | Homo sapi   |
| C 12       | 419.6 | 41.9        | 167717 | 9  | AL390198 | Human DNA   |
| C 13       | 330.6 | 33.0        | 178105 | 9  | AC073057 | Homo sapi   |
| C 14       | 327.8 | 32.7        | 140653 | 2  | AL355134 | Homo sapi   |
| C 15       | 327.8 | 32.7        | 150286 | 2  | AL353644 | Homo sapi   |
| C 16       | 326.2 | 32.6        | 126187 | 2  | AL592188 | Homo sapi   |
| C 17       | 306.4 | 30.6        | 57476  | 2  | AC107630 | Homo sapi   |
| C 18       | 292.8 | 29.3        | 175952 | 9  | AC069285 | Homo sapi   |
| C 19       | 264.6 | 26.4        | 139384 | 9  | AC006455 | Homo sapi   |
| C 20       | 263   | 26.3        | 134187 | 9  | AC006457 | Homo sapi   |
| C 21       | 254.6 | 25.4        | 64824  | 2  | AC067782 | Homo sapi   |
| C 22       | 254.4 | 25.4        | 167133 | 2  | AC026186 | Homo sapi   |
| C 23       | 253.2 | 25.3        | 178105 | 9  | AC073057 | Homo sapi   |
| C 24       | 244   | 24.4        | 208825 | 2  | AC119407 | Pan trogl   |
| C 25       | 238   | 23.8        | 104660 | 9  | AP000347 | Homo sapi   |
| C 26       | 235   | 23.5        | 677    | 9  | HS329573 | Homo sapi   |
| C 27       | 220.4 | 22.0        | 862    | 9  | HS342340 | Homo sapi   |
| C 28       | 199.4 | 19.9        | 637    | 9  | HS338826 | Homo sapi   |
| C 29       | 183.4 | 18.3        | 165432 | 9  | AL512590 | Human DNA   |
| C 30       | 183.4 | 18.3        | 194142 | 9  | AC092634 | Homo sapi   |
| C 31       | 181.8 | 18.2        | 167133 | 2  | AC026186 | Homo sapi   |
| C 32       | 166   | 16.6        | 603    | 9  | HS342444 | Homo sapi   |
| C 33       | 132.8 | 13.3        | 115916 | 9  | AC115220 | Homo sapi   |
| C 34       | 132   | 13.2        | 129290 | 2  | AC020761 | Homo sapi   |
| C 35       | 123   | 12.3        | 725    | 9  | HS339687 | Homo sapi   |
| C 36       | 121.8 | 12.2        | 676    | 9  | HS329644 | Homo sapi   |
| C 37       | 121.4 | 12.1        | 660    | 9  | HS339184 | Homo sapi   |
| C 38       | 117.6 | 11.7        | 125419 | 2  | AC026260 | Homo sapi   |
| C 39       | 115.6 | 11.5        | 146383 | 2  | AC016052 | Homo sapi   |
| C 40       | 115   | 11.5        | 188679 | 9  | AC073270 | Homo sapi   |
| C 41       | 112.2 | 11.2        | 662    | 9  | HS338698 | Homo sapi   |
| C 42       | 111.4 | 11.1        | 215287 | 2  | AC079432 | Mus muscu   |
| C 43       | 111   | 11.1        | 135351 | 9  | AC068279 | Homo sapi   |
| C 44       | 111   | 11.1        | 164485 | 2  | AC026106 | Homo sapi   |
| C 45       | 111   | 11.1        | 167220 | 2  | AC013411 | Homo sapi   |

ALIGNMENTS

RESULT 1  
AC018558

LOCUS

DEFINITION

ACCESSION

VERSION

KEYWORDS

SOURCE

ORGANISM

REFERENCE

AUTHORS

TITLE

JOURNAL

AC018558 Homo sapiens chromosome 16 clone RP11-80F22, complete sequence.  
151049 bp DNA linear PRI 02-NOV-2001

AC018558

AC018558

AC018558.5 GI:16596530

HTG.

Homo sapiens.

Homo sapiens

Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;

Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.

1 (bases 1 to 151049)

DOE Joint Genome Institute.

Sequencing of Human Chromosome 16

Unpublished

REFERENCE 2 (bases 1 to 151049)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (14-DEC-1999) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

REFERENCE 3 (bases 1 to 151049)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (02-NOV-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA

COMMENT On Nov 2, 2001 this sequence version replaced gi:795566.  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

Sequence Quality Assessment:  
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.  
Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.

FEATURES  
source Location/Qualifiers  
1. 151049  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="16"  
/clone="RP11-80F22"  
BASE COUNT 46827 a 30783 c 30749 g 42690 t  
ORIGIN

Query Match 100.0%; Score 1001; DB 9; Length 151049;  
Best Local Similarity 100.0%; Pred. No. 5.2e-304; Mismatches 0; Indels 0; Gaps 0;  
Matches 1001; Conservative 0;

QY 1 AGCAACCTGTAGTTCGGGCTGCAATCATAGATAAGTAGGAGCTTGTATGGGAG 60  
DB 118023 AGCAACCTGTAGTTCGGGCTGCAATCATAGATAAGTAGGAGCTTGTATGGGAG 118082

QY 61 GGATGGCTGCAGCTTCATGGATAGAAATGTCAGCTTGGGCTAGATACCAACATGGG 120  
DB 118083 GGATGGCTGCAGCTTCATGGATAGAAATGTCAGCTTGGGCTAGATACCAACATGGG 118142

QY 121 GCCTCACTCCTCTTTGTAGCACACGCCACCATAGAAAGAGATAGCACTTGGATAGC 180  
DB 118143 GCCTCACTCCTCTTTGTAGCACACGCCACCATAGAAAGAGATAGCACTTGGATAGC 118202

QY 181 TCAAAAGTACAGGAGCTTCAGTGTCCCTTCTGTGGAGCCAGCAACTGATGAGCTCTAA 240  
DB 118203 TCAAAAGTACAGGAGCTTCAGTGTCCCTTCTGTGGAGCCAGCAACTGATGAGCTCTAA 118262

QY 241 GTCCTGTTGTATGAACATGTCTCTACCCCTGGGCGCCCTGGTGGTGCAGCATAGGAAG 300  
DB 118263 GTCCTGTTGTATGAACATGTCTCTACCCCTGGGCGCCCTGGTGGTGCAGCATAGGAAG 118322

QY 301 TATAAGGATAGGCTTAGTCTAGTCATGGCCATGAGCCTTTCATTAACTTGGCTGCTG 360  
DB 118323 TATAAGGATAGGCTTAGTCTAGTCATGGCCATGAGCCTTTCATTAACTTGGCTGCTG 118382

QY 361 CCTTCTAGGGAATATAATCAACACTAAATAAGGAGGAGGCTGAGCAGCTGGCGCTGTCG 420  
DB 118383 CCTTCTAGGGAATATAATCAACACTAAATAAGGAGGAGGCTGAGCAGCTGGCGCTGTCG 118442

QY 421 TTTGAGGAGGATGGCGATGTGAAGTCAAGTCAACCCCTGGGAGGACACTCCCTGGCT 480

Db 118443 TTTGAGGAGGATGGCGATGTGAAGTCAGTCAACCCGTTGGGAGGACACTCCCTGGCT 118502  
QY 481 CCATCCTCTGCATCTTAGATTATTATGGGACAGTATTGATACACAGAGAAGGAGAGACCCA 540  
Db 118503 CCATCCTCTGCATCTTAGATTATTATGGGACAGTATTGATACACAGAGAAGGAGAGACCCA 118562  
QY 541 TCCCAATGGAGGGTTTGTATAGATGAATATATCAATCAATGAATAATCTCTAGAGGAGGACT 600  
Db 118563 TCCCAATGGAGGGTTTGTATAGATGAATATATCAATCAATGAATAATCTCTAGAGGAGGACT 118622  
QY 601 TTTTATATCAACTCTGAGAACAGTTGGAGCTACATGGGATTGGAGGGAGGCTGGAGC 660  
Db 118623 TTTTATATCAACTCTGAGAACAGTTGGAGCTACATGGGATTGGAGGGAGGCTGGAGC 118682  
QY 661 CCCTTAAAGAAAAAGCCCCAGAGACTGCCCTGCCCTCTCTCTCCCCACAAAGTTCCATT 720  
Db 118683 CCCTTAAAGAAAAAGCCCCAGAGACTGCCCTGCCCTCTCTCTCCCCACAAAGTTCCATT 118742  
QY 721 TATTATCTTCCACCCAGGAGCTGTGAGAAATCCTGCCCTTCCGTTCCAGATCAAAAGTCT 780  
Db 118743 TATTATCTTCCACCCAGGAGCTGTGAGAAATCCTGCCCTTCCGTTCCAGATCAAAAGTCT 118802  
QY 781 TCAGAAATGCAACTACTTTCAGTCACAGAGATAATTTATCATCTTCTGACAGAGAGGAA 840  
Db 118803 TCAGAAATGCAACTACTTTCAGTCACAGAGATAATTTATCATCTTCTGACAGAGAGGAA 118862  
QY 841 TTTGGGTTTGGTCCCGTCCAGTCCATGAAGTGGCACAGTCAAGATAAAGGTGAGAGCTTAGG 900  
Db 118863 TTTGGGTTTGGTCCCGTCCAGTCCATGAAGTGGCACAGTCAAGATAAAGGTGAGAGCTTAGG 118922  
QY 901 AGATTAGCGGAGGTAGAGAACACTCTGTCTTTGTGACCACTTCAGAGAGCCCTGGGGCC 960  
Db 118923 AGATTAGCGGAGGTAGAGAACACTCTGTCTTTGTGACCACTTCAGAGAGCCCTGGGGCC 118982  
QY 961 ATGGCTTCTGCTCAACATTAGGCCCTGCTGCATGGTGACC 1001  
Db 118983 ATGGCTTCTGCTCAACATTAGGCCCTGCTGCATGGTGACC 119023

RESULT 2  
AC092357 AC092357 191768 bp DNA linear PRI 15-AUG-2001  
LOCUS Homo sapiens chromosome 16 clone RP11-332P24, complete sequence.  
DEFINITION AC092357 AC092357  
ACCESSION AC092357  
VERSION AC092357.2 GI:15187270  
KEYWORDS HTG.  
SOURCE Homo sapiens.  
ORGANISM Homo sapiens  
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
REFERENCE 1 (bases 1 to 191768)  
AUTHORS DOE Joint Genome Institute.  
TITLE Sequencing of Human Chromosome 16  
JOURNAL Unpublished  
REFERENCE 2 (bases 1 to 191768)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (03-JUL-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
REFERENCE 3 (bases 1 to 191768)  
AUTHORS DOE Joint Genome Institute.  
TITLE Direct Submission  
JOURNAL Submitted (15-AUG-2001) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
COMMENT On or before Aug 15, 2001 this sequence version replaced gi:7596817, gi:14589546.  
Sequence Quality Assessment:  
This entry has been annotated with sequence quality estimates computed by the Phrap assembly program.  
All manually edited bases have been reduced to quality zero.  
Quality levels above 40 are expected to have less than 1 error in 10,000 bp.



|                       |   |   |              |           |            |     |             |             |
|-----------------------|---|---|--------------|-----------|------------|-----|-------------|-------------|
| Query Match           |   | 65.9%;  | Score        | 659.2;    | DB         | 9;  | Length      | 191014;     |
| Best Local Similarity |   | 89.2%;  | Pred. No.    | 4.3e-196; |            |     |             |             |
| Matches               |   | 756;  | Conservative | 0;        | Mismatches | 63; | Indels      | 29; Gaps 3; |
| QY                    | 183   | AAAAGTCACGGAGCCTCAGTGTCCCTTCTCTGTGAGGCCACGAACCTGATGACAGGTCTAAGT | 242          |           |            |     |             |             |
| DB                    | 153394  | AAAGTTCACGGAGCCTCAGTGTCCCTTCTCTGTGAGGCCACGAACCTGATGACAGGTCCAAGT | 153335       |           |            |     |             |             |
| QY                    | 243   | CCTGTTGTATGAACATGTCTGACCTTGGCGGCCCTGGTGGTGGTGCAGCATAGGAAGTA     | 302          |           |            |     |             |             |
| DB                    | 153334  | CCTATTGTATGAACATGTCTGACCTTTCGGCCCTGCTGGTGGTGCAGCTAGGAAGTA       | 153275       |           |            |     |             |             |
| QY                    | 303   | TAAGGGATGAGGTCTAGTCATGGGCCATGGAGCCCTTCTCATTAATCTTGGCTGCTGACC    | 362          |           |            |     |             |             |
| DB                    | 153274  | TAAGGGACGAGTCTAGTCATGGGCCACGGAGCCCTTCTCATGAATCTTGGCTGCTTACC     | 153215       |           |            |     |             |             |
| QY                    | 363   | TTCTAGG-GAATATAATCAACACTAATAAAGGAGGAGGTGCAGCAGCTGGCGCTGTCGCT    | 421          |           |            |     |             |             |
| DB                    | 153214  | TTCTAGGAAAATAATCAACACTAATAAAGCGGAAGGTGAGCAGCGGCACCTGTGCTT       | 153155       |           |            |     |             |             |
| QY                    | 422   | TTGAGGAGGATGGCGATGTGAAAGTCAGTGACCAACCGTGGGGAGGACACTCCCTGGCTC    | 481          |           |            |     |             |             |
| DB                    | 153154  | TTGAGGAGGATGGCGATGTGACAGTCAGAGCACCAACCGTGGGGAGGACACTCCCTGACTC   | 153095       |           |            |     |             |             |
| QY                    | 482   | CATCCTCTGCATCTTAGATTTATTGGGACAGTTTGATACACAGAG-----              | 526          |           |            |     |             |             |
| DB                    | 153094  | CATCCTCTGCATCTTAGATTTACTGGGACGGTTTGATTAGAGGAGTGTGTTGATTAGAGGA   | 153035       |           |            |     |             |             |
| QY                    | 527   | -----AAGGAGGAGACCCATCCCAATGGAGGTTTGATTAGATCAATATATCAAT          | 577          |           |            |     |             |             |
| DB                    | 153034  | TTTGGTCCAAAGTCTGAGCCAAAGATCAGTGGAGGGTTTGATTAGAGGAATAGAAACAAT    | 152975       |           |            |     |             |             |
| QY                    | 578   | GATAAATCTCTAGAGGAGGACTTTTATAATCAACTCTGAGAACAGGTTGGAGCTACAT      | 637          |           |            |     |             |             |
| DB                    | 152974  | GATAAATCTCTAGAGGAGGACTGTTTATAATCAACTCTGAGAACAGGTTGGGGCTACAT     | 152915       |           |            |     |             |             |
| QY                    | 638   | GGGATTGGAGGGAGGGTGGAGCCCTTAAAGAAAAGCCCCAGAGACTGCCCTGCGCCT       | 697          |           |            |     |             |             |
| DB                    | 152914  | GGGATTGGAGGGAGGGTGGAGCCCTTAAAGAAAAGCTCCAGAGACTGCCCTGCGCCT       | 152855       |           |            |     |             |             |
| QY                    | 698   | CTCTCTCCCCACAAGTTCCATTTTATTTATCTTCCACCCAGAGCTGTGAGAACTCTGCC     | 757          |           |            |     |             |             |
| DB                    | 152854  | CTCCCTCCCCGACAAGTCTCTTTTATTATCTTCCACCCAGGAGTGTGAGAACTCTGCC      | 152795       |           |            |     |             |             |
| QY                    | 758   | TTCC-----GTCTCCAGATCAAAAGTCTTCAGGAAATGCAACTACTTCAGTGACAAGAGAT   | 813          |           |            |     |             |             |
| DB                    | 152794  | TTCTTCTGTCTCCAGATCAAAAGTCTTCAGGAAATGCAATCTGCTTCAAGTCAACAAGAT    | 152735       |           |            |     |             |             |
| QY                    | 814   | AATTATCATCTCTGCAGAGGAGGAATTTGGGGTTTGGTCCCAAGTCCATGAAGTGGCAC     | 873          |           |            |     |             |             |
| DB                    | 152734  | AATTATCATCTCTGCAGAGGAGGAATTTGGGGTTTGGTCCCAAGTCCATGAAGTGGCAC     | 152675       |           |            |     |             |             |
| QY                    | 874   | AGTCAGAAATAAAGGTTCAGAGCTTAGGAGATTACGGGGGTAGAGAACACTCTGTCTT      | 933          |           |            |     |             |             |
| DB                    | 152674  | AGTCAGAAATAAAGGTTCAGGGCCTTAGGAGATTAGCAGAGGTTAGAGAAGACTCTATCTT   | 152615       |           |            |     |             |             |
| QY                    | 934   | GTGACCAAGCTTCAGAGAGCCTGGGGCCATGGCTTCTTGTCGAACATTAGGCCCTGCTGCA   | 993          |           |            |     |             |             |
| DB                    | 152614  | CTGACCAAGCTTCAGAGAGCCTGGAGCCATGGCTTCTTGTCGAACATTAGGCCCTGCTGCA   | 152555       |           |            |     |             |             |
| QY                    | 994   | TGCTGACC 1001   |              |           |            |     |             |             |
| DB                    | 152554  | TGGGGACC 152547   |              |           |            |     |             |             |
| RESULT 4              |   |   |              |           |            |     |             |             |
| LOCUS                 | AC007882  | 214025 bp   |              | DNA       | linear     | PRI | 22-MAY-2002 |             |
| DEFINITION            | Homo sapiens BAC clone RP11-499D5 from 16, complete sequence. |   |              |           |            |     |             |             |
| ACCESSION             | AC007882  |   |              |           |            |     |             |             |
| VERSION               | AC007882.3  |   |              |           |            |     |             |             |
| KEYWORDS              | HTG.  |   |              |           |            |     |             |             |
| SOURCE                | Homo sapiens.   |   |              |           |            |     |             |             |

|           |  |  |
|-----------|--|--|
| ORGANISM  |  | Homo sapiens   |
| REFERENCE |  | Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.   |
| AUTHORS   |  | 1 (bases 1 to 214025)  |
| TITLE     |  | Sulston,J.E. and Waterston,R.  |
| JOURNAL   |  | Toward a complete human genome sequence  |
| MEDLINE   |  | Genome Res. 8 (11), 1097-1108 (1998)   |
| PUBMED    |  | 99063792   |
| REFERENCE |  | 9847074  |
| AUTHORS   |  | 2 (bases 1 to 214025)  |
| TITLE     |  | Reitz,L., Stoneking,T. and Andrews,S.  |
| JOURNAL   |  | The sequence of Homo sapiens BAC clone RP11-499D5  |
| REFERENCE |  | Unpublished (2001)   |
| AUTHORS   |  | 3 (bases 1 to 214025)  |
| TITLE     |  | Waterston,R.H.   |
| JOURNAL   |  | Direct Submission  |
| REFERENCE |  | Submitted (19-JUN-1999)  |
| AUTHORS   |  | Genome Sequencing Center, Washington   |
| TITLE     |  | University School of Medicine, 4444 Forest Park Parkway, St. Louis,  |
| JOURNAL   |  | MO 63108, USA  |
| REFERENCE |  | 4 (bases 1 to 214025)  |
| AUTHORS   |  | Waterston,R.H.   |
| TITLE     |  | Direct Submission  |
| JOURNAL   |  | Submitted (20-APR-2000)  |
| REFERENCE |  | Genome Sequencing Center, Washington   |
| AUTHORS   |  | University School of Medicine, 4444 Forest Park Parkway, St. Louis,  |
| TITLE     |  | MO 63108, USA  |
| JOURNAL   |  | 5 (bases 1 to 214025)  |
| REFERENCE |  | Waterston,R.H.   |
| AUTHORS   |  | Direct Submission  |
| TITLE     |  | Submitted (21-APR-2000)  |
| JOURNAL   |  | Genome Sequencing Center, Washington   |
| REFERENCE |  | University School of Medicine, 4444 Forest Park Parkway, St. Louis,  |
| AUTHORS   |  | MO 63108, USA  |
| TITLE     |  | 6 (bases 1 to 214025)  |
| JOURNAL   |  | Waterston,R.H.   |
| REFERENCE |  | Direct Submission  |
| AUTHORS   |  | Submitted (22-APR-2000)  |
| TITLE     |  | Genome Sequencing Center, Washington   |
| JOURNAL   |  | University School of Medicine, 4444 Forest Park Parkway, St. Louis,  |
| REFERENCE |  | MO 63108, USA  |
| AUTHORS   |  | 7 (bases 1 to 214025)  |
| TITLE     |  | Waterston,R.   |
| JOURNAL   |  | Direct Submission  |
| REFERENCE |  | Submitted (30-SEP-2000)  |
| AUTHORS   |  | Department of Genetics, Washington   |
| TITLE     |  | University, 4444 Forest Park Avenue, St. Louis, Missouri 63108, USA  |
| JOURNAL   |  | On Apr 20, 2000 this sequence version replaced gi:5836167.   |
| COMMENT   |  | ----- Genome Center<br>Center: Washington University Genome Sequencing Center<br>Center code: WUGSC<br>Web site: http://genome.wustl.edu/gsc<br>Contact: sapiens@watson.wustl.edu<br>----- Summary Statistics<br>Center project name: H_NH0499D05<br>----- |

NOTICE: This sequence may not represent the entire insert of this clone. It may be shorter because we only sequence overlapping clone sections once, or longer because we provide a small overlap between neighboring data submissions.

This sequence was finished as follows unless otherwise noted: all regions were double stranded, sequenced with an alternate chemistry, or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by sequence from more than one subclone; and the assembly was confirmed by restriction digest.

MAPPING INFORMATION:  
Mapping information for this clone was provided by Dr. John D. McPherson, Department of Genetics, Washington University, St. Louis



```
Qy 482 CATCTCTGCATCTTAGATTTATTGGGACATGTTGTATACACAGAGAGAGGAGGCCCAT 541
    |||||
Db 29768 CATCTCTGCATCTTAGATTTATTGGGAAGCTTTGTATACACAGAGAGAGGAGGCCCAT 29827
Qy 542 CCCAATGGAGGTTTGATAGATCAATATATATCAATGATTAATTCCTTAGAGGAGGGACTT 601
    |||||
Db 29828 CCCAATGGAGGTTTGATAGGGAATAGAAATCAATGATAAATCCCTTAGAGGAGGGACTG 29887
Qy 602 TTTATATCAAACTC-TGAGAACAGGTTGGAGCTACATGGGATTTGAGGGAGGGTGGAGC 660
    |||||
Db 29888 TTTATATCAAACTCTTGAGAACAGGTTGGGGCTACATGGGATTTGAGGGAGGGTGGAGC 29947
Qy 661 CCCTTAAAGAAAAGCCCCAGAGAGTGCCTCGCCCTCTCTCTCCCCACAAAGTTCCATTT 720
    |||||
Db 29948 CCCTTAAAGAAAAGCCCCTAGAAAATGTGCCCTTACACCCCCCACAAAGTTCCCTT 30007
Qy 721 TATTATCTTCCACCCAGAGCTGTGCAGATCCTGCCCTTCC----GTCCTCCAGATCAAG 776
    |||||
Db 30008 TGTCAATCTTCCACCCAGAGCTGTGCAGAAATCCTGCCCTTCTCTGTCTCCAGATCAAG 30067
Qy 777 TCCTTCAGGAAATGCACTACTTTCAGTGACAAGAGATAATATCATCTTCTGCAGACAGGA 836
    |||||
Db 30068 TCCTCCAGGAAATGCAGCTGCTTCAGTGACAAGAGATAATGTCTATCTTCTGACTGAGGA 30127
Qy 837 GGAATTTGGGTTGGTCCCACTCATGAAGTGGCAGAGTCAAGATAAAAGTGGAGAGCT 896
    |||||
Db 30128 GGAATTTGGGTTGGTTCAGTGCATGAAGCTGCACAGTCAAGATAAAAGTGGAGAGCC 30187
Qy 897 TAGGAGATTACGGAGGTAGAGAACACTCTGCTTGTGACCACTGCAGAGTCCAGAGCCCTGG 956
    |||||
Db 30188 TAGCAGATTAGCAAGACTAGGAGAAGACTCTATCTTGTGGCCAGCTTCAGAGAACCTGG 30247
Qy 957 GGCATAGTCTCCTGGTCAACATTAGGCCCTGCTGCATGGTGACC 1001
    |||||
Db 30248 GGCATAGTCTCCTGGTCAACATTAGG-CCTGCTGCATGGGGACC 30291

RESULT 5
AC007908/c
LOCUS      227856 bp      DNA      linear      HTG 25-APR-2001
DEFINITION Homo sapiens chromosome 16 clone RP11-499D5, WORKING DRAFT
SEQUENCE, 30 unordered pieces.
AC007908
VERSION    AC007908.3  GI:9094205
KEYWORDS   HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE     Homo sapiens.
ORGANISM   Homo sapiens
            Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
REFERENCE  1 (bases 1 to 227856)
            DOE Joint Genome Institute.
            Sequencing of Human Chromosome 16
            Unpublished
            2 (bases 1 to 227856)
            Bruce,D., Mundt,M., Doggett,N., Munk,C., Saunders,E., Robinson,D.,
            Jones,M., Buckingham,J., Chasteen,L., Thompson,S., Goodwin,L.,
            Bryant,J., Tesmer,J., Meincke,L., Longmire,J., White,S., Tatum,O.,
            Campbell,C., Fawcett,J., Malthie,M., Bussod,M., Sutherland,R.,
            McMurry,K., Han,d. and Deaven,L.
            Direct Submission
            Submitted (24-JUN-1999) Center for Human Genome Studies, DOE Joint
            Genome Institute, Los Alamos National Laboratory, MS M888, Los
            Alamos, NM 87545, USA
            On Jul 13, 2000 this sequence version replaced gi:7211873.
COMMENT    -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 0
            Center clone name: RPCI-11_499D5
            -----
```

```
Summary Statistics
Consensus quality: 184734 bases at least Q40
Consensus quality: 201642 bases at least Q30
Consensus quality: 209314 bases at least Q20
Estimated insert size: 218590; agarose-fp estimation
Estimated insert size: 224956; sum-of-contigs estimation
Quality coverage: 4.43 in Q20 bases; agarose-fp estimation
Quality coverage: 4.43 in Q20 bases; sum-of-contigs estimation.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 30 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

```
* 1 1259: contig of 1259 bp in length
* 1359: gap of unknown length
* 2455: contig of 1096 bp in length
* 2555: gap of unknown length
* 3691: contig of 1136 bp in length
* 3791: gap of unknown length
* 4805: contig of 1014 bp in length
* 4906: gap of unknown length
* 5938: contig of 1033 bp in length
* 6038: gap of unknown length
* 7051: contig of 1013 bp in length
* 7152: gap of unknown length
* 8246: contig of 1095 bp in length
* 8346: gap of unknown length
* 9993: contig of 1647 bp in length
* 10093: gap of unknown length
* 11375: contig of 1282 bp in length
* 11475: gap of unknown length
* 12513: contig of 1038 bp in length
* 12613: gap of unknown length
* 13645: contig of 1032 bp in length
* 13745: gap of unknown length
* 15252: contig of 1507 bp in length
* 15352: gap of unknown length
* 16573: contig of 1221 bp in length
* 16673: gap of unknown length
* 18205: contig of 1532 bp in length
* 18305: gap of unknown length
* 19913: contig of 1608 bp in length
* 20013: gap of unknown length
* 22479: contig of 2466 bp in length
* 22579: gap of unknown length
* 25293: contig of 2714 bp in length
* 25393: gap of unknown length
* 29118: contig of 3725 bp in length
* 32189: gap of unknown length
* 32189: contig of 2971 bp in length
* 32289: gap of unknown length
* 33294: contig of 1005 bp in length
* 33394: gap of unknown length
* 34509: contig of 1115 bp in length
* 34609: gap of unknown length
* 36147: contig of 1538 bp in length
* 36247: gap of unknown length
* 42248: contig of 6001 bp in length
* 43348: gap of unknown length
* 52325: contig of 9977 bp in length
* 52425: gap of unknown length
* 60479: contig of 8054 bp in length
* 60579: gap of unknown length
* 68636: gap of unknown length
* 68737: contig of 8057 bp in length
* 83105: contig of 14369 bp in length
* 83205: gap of unknown length
* 116897: contig of 33692 bp in length
* 116997: gap of unknown length
* 165561: contig of 48564 bp in length
```



|   |  |
|---|--|
| * 165562 165661: gap of unknown length                        |  |
| * 165662 227856: contig of 62195 bp in length.                |  |
| FEATURES  | Location/Qualifiers  |
| source  | 1..227856  |
|   | /organism="Homo sapiens"   |
|   | /db_xref="taxon:9606"  |
|   | /chromosome="16"   |
|   | /clone_lib="RPCI human BAC library 11"                                       |
| BASE COUNT  | 68439 a 46078 c 44943 g 64859 t 3537 others                                  |
| ORIGIN  |  |
| Query Match 64.6%; Score 646.2; DB 2; Length 227856;          |  |
| Best Local Similarity 89.7%; Pred. No. 5.5e-192;              |  |
| Matches 740; Conservative 0; Mismatches 78; Indels 7; Gaps 4; |  |
| QY  | 183 AAAAGTCAGGAGCCTCAGCTGCCCTTCGTGCGGCCAGACCTGATGCAGGCTTAAGT 242             |
| Db  | 200340 AAAGGTCATGAGCCCTCAGTGTCCCTTCGTGTGGGCCAGAACCTGATGCAGGCTTAAGT 200281    |
| QY  | 243 CCTGTTGATGAACATGTCTCCTGACCCCTGGCGCCCTGTGTGGTGCAGCATAGGAAGTA 302          |
| Db  | 200280 CCGTTTTATGCACATGCTCTGACCTGTGTGCCCTGTGTGGTGCAGCATAGGAAGTA 200221       |
| QY  | 303 TAAGGATGAGGTCCTAGTCATGGCGCATGGAGCCCTTCTCATTAATCTTGGCTGCTGCC 362          |
| Db  | 200220 TAAGGATGAGGTCCTAGTCCTGTGTGAGGAAGCCCTTCTCATGAATCTTGGCTGCTACC 200161    |
| QY  | 363 TTCTAGGGA-ATATATCAACACTAATAAGAGAGAGAGGTGAGCAGCTGGCGCTGTCCT 421           |
| Db  | 200160 TCTTAAGACATATATCAACACTAATAAGGAAGAGGTGAGCAGCTGGCGCTGTCCT 200101        |
| QY  | 422 TTGAGGAGGATGGCGATGGAAGTCAGTGACACCGCTGGGAGGACACTCCCTGGCTC 481             |
| Db  | 200100 TTGAGGAGGATGGCGATGGAAGTCAGTGACACACCCCTGAGGAAGACACTCCTTGGCTC 200041    |
| QY  | 482 CATCTCTGCATCTTAGATTATTGGAAGTTTGTATACAGAGAAGGAGAGACCCAT 541               |
| Db  | 200040 CATCTCTGCATCTTAGATTATTGGAAGTTTGTATACAGAGAAGGAGAGACCCAT 199981         |
| QY  | 542 CCCAATGGAGGTTTCATTAGATGAATATAATCAATGAATATTCCTAGAGAGGAGCTT 601            |
| Db  | 199980 CCCAATGGAGGTTTCATTAGAGGAATGAATCAATGATAAATCCTAGAGAGGAGCTG 199921       |
| QY  | 602 TTTTATATCAACTC-TGAGAACAGGTTGGAGTCATCGGATTTGGAGGGAGGTTGGAGC 660           |
| Db  | 199920 TTTTATATCAACTCTTGAGAACAGGTTGGGCTACATGGATTTGGAGGGAGGTTAGAC 199861      |
| QY  | 661 CCCTTAAAGAAAGCCCGCAGAGATGCCCTTGCCCTCTCTCTCCCCACAAAGTTCCATT 720           |
| Db  | 199860 CCCTTAAAGAAAGCCCTTAGAAATTTGCCCTTACCCCTTCACACCCCCACAAAGTTCCCTT 199801  |
| QY  | 721 TATTATCTTCCACCCAGGAGCTGACAGATCTGCGCTCCCTTCC- - -GTCTCCAGATCAAG 776       |
| Db  | 199800 TGTATCTTCCACCCAGGAGCTGACAGATCTGCGCTCCCTTCTGTCTCCAGATCAAG 199741       |
| QY  | 777 TCCTTCAGGAAATGCAACTACTCTAGTCACAAAGAGATAATTATCATCTTCTGCACAGAGGA 836       |
| Db  | 199740 TCCTTCAGGAAATGCAAGCTGCTTCAAGTCACAGAGATAATTCTCATCTTCTGACTGAGGA 199681  |
| QY  | 837 GGAATTTGGGTTTGTGTCCTCATGTCATGAAGTGGCAGAGTCAGATAAAAGGTGAGAGCT 896         |
| Db  | 199680 GGAATTTGGGTTTGTGTCCTCATGTCATGAAGTGGCAGAGTCAGATAAAAGATGAGACC 199621    |
| QY  | 897 TAGGACATTAGCGGAGGTGAGAAGACACTCTGCTTGTGACCACTTCAGAGAGCCCTGG 956           |
| Db  | 199620 TAGGACATTAGCAAGAGACTAGAGAAGACTCTATCTTGTGCGCCAGCTTCAGAGAACCCTGG 199561 |
| QY  | 957 GGCCATGGCTTCTGCTCAACATTAGGCCCTCTGCTCATGTTGACC 1001                       |
| Db  | 199560 GGCCATAGTCTCCTGCTCAACATTAGG-CTTGCTCATGTTGGGACC 199517                 |





```
Db 67192 AAAGGTCACTAGACCTCAGTGTCCCTTCTGTGGAGCCAGAACCTCATGACGTCTAAGT 67133
Qy 243 CTTGTTGTATGAACATGTCTCAGCCTGCGGCCCTGGTGGTGGTGCAGCATAGGAAGTA 302
Db 67132 CTTGTTTATGACATGTCTCAGCCTGCGGCCCTGGTGGTGGTGCAGCATAGGAAGTA 67073
Qy 303 TAAGGATGAGGTCTAGTTCATGGGCATGGAGCCTTCTCATTAATCTTGGCTGTCTGCC 362
Db 67072 TAAGGATGAGGTCTAGTTCATGGGCATGGAGCCTTCTCATTAATCTTGGCTGTCTGCC 67013
Qy 363 TTCTAGGGA-ATATAATCAACACTAATAAAGGAGGAGGTGAGAGCTGGCGCTGTGCGT 421
Db 67012 TCTTAAGAACATATAATCAACACTAATAAAGGAGGAGGTGAGAGCTGGCGCTGTGCGT 66953
Qy 422 TTGAGGAGGATGGCATGTGAAGTCACTGACCCCGTGGGAGGAGACACTCCCTGGCTC 481
Db 66952 TTGAGGAGGATGGCATGTGAAGTCACTGACCCCGTGGGAGGAGACACTCCCTGGCTC 66893
Qy 482 CATCTCTGCATCTTAGATTTATTTGGGACAGTTTATGATACACAGAGAGGAGAGCCCAT 541
Db 66892 CATCTCTGCATCTTAGATTTATTTGGGAGGTTTATGATACACAGAGAGGAGAGCCCAT 66833
Qy 542 CCATGAGGAGGTTGATTTAGATGAATATAATCAATGATAATTTCTTAGAGGAGGACTT 601
Db 66832 CCATGAGGAGGTTGATTTAGGGAATAGATAATCAATGATAATTTCTTAGAGGAGGACTT 66773
Qy 602 TTTATATCAACTC-TGACACAGGTTTGAGGTACATGAGTGGATGGAGGAGGAGGAGC 660
Db 66772 TTTATATCAACTC-TGACACAGGTTTGAGGTACATGAGTGGATGGAGGAGGAGGAGC 66713
Qy 661 CCCTTAAAGAAAGCCCGACAGACTGCCCCCTGCTCTCTCCGCCACAAAGTTCCAPT 720
Db 66712 CCCTTAAAGAAAGCCCGTAGAATTTGCTCCCTGACCCCTTCACACCCCGACAAAGTTCCAPT 66653
Qy 721 TATTATCTCCACCCAGGAGCTGTGAGATCTGCGCCCTTCC- - -GTCTCCAGATCAAAAG 776
Db 66652 TGTCACTCTCCACCCAGGAGCTGTGAGATCTGCGCCCTTCCCTGCTCCAGATCAAAAG 66593
Qy 777 TCTCTCAGGAATGCAACTACTCTCAGTGACAGAGATTAATCATCTCTCAGAGGA 836
Db 66592 TCTCTCAGGAATGCAACTACTCTCAGTGACAGAGATTAATCATCTCTCAGAGGA 66533
Qy 837 GGAATTTGGGGTTGGTCCAGTCCATGAAGTGGCAGCTCAGATCAAGATTAAGGTGAGAGCT 896
Db 66532 GGAATTTGGGGTTGGTCCAGTCCATGAAGTGGCAGCTCAGATCAAGATTAAGGTGAGAGCT 66473
Qy 897 TAGGAGATTAGCGGAGGTGAGAGACACTGTCTTGTGACAGGCTTCAGAGAGGCTGG 956
Db 66472 TAGCAGATTAGCAAGACTAGGAGAGACTCTATCTTGTGGCCAGCTTCAGAGAGGCTGG 66413
Qy 957 GGCATGCTTCTCTGCTCAACATTAGGCCCTGCTGCTGCTGAC 1001
Db 66412 GGCATGCTTCTCTGCTCAACATTAGG-CTGCTGCTGCTGAGGACC 66369

RESULT 8
AC114938/c
LOCUS      167553 bp      DNA      linear      HTG 14-MAR-2002
DEFINITION Homo sapiens chromosome 5 clone CTD-2244D9, WORKING DRAFT SEQUENCE,
            3 unordered pieces.
ACCESSION  AC114938
VERSION    AC114938.1  GI:19424399
KEYWORDS  HTG; HTGS_PHASE1; HTGS_DRAFT; HTGS_ACTIVEFIN.
SOURCE    Homo sapiens.
ORGANISM  Homo sapiens
REFERENCE Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
            Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
            1 (bases 1 to 167553)
AUTHORS   DOE Joint Genome Institute.
TITLE     Sequencing of Human Chromosome 5
JOURNAL   Unpublished
REFERENCE 2 (bases 1 to 167553)
```

```
AUTHORS   DOE Joint Genome Institute.
TITLE     Direct Submission
JOURNAL   Submitted (14-MAR-2002) Production Sequencing Facility, DOE Joint
            Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA
COMMENT   -----Genome Center
            Center: Joint Genome Institute
            Center Code: JGI
            Web site: http://www.jgi.doe.gov
            -----
            Project Information
            Center Project Name: 721326
            Center clone name: CITB-HI_2244D9
            -----
            Summary Statistics
            Consensus quality: 163638 bases at least Q40
            Consensus quality: 165018 bases at least Q30
            Consensus quality: 166006 bases at least Q20
            Estimated insert size: 150000; agarose-fp estimation
            Estimated insert size: 167353; sum-of-contrigs estimation
            Quality coverage: 10.56 in Q20 bases; agarose-fp estimation
            Quality coverage: 9.46 in Q20 bases; sum-of-contrigs estimation.
            * NOTE: This is a 'working draft' sequence. It currently
            * consists of 3 contigs. The true order of the pieces
            * is not known and their order in this sequence record is
            * arbitrary. Gaps between the contigs are represented as
            * runs of N, but the exact sizes of the gaps are unknown.
            * This record will be updated with the finished sequence
            * as soon as it is available and the accession number will
            * be preserved.
            *
            * 1 31475: contig of 31475 bp in length
            * 31476 31575: gap of unknown length
            * 31576 82160: contig of 50585 bp in length
            * 82161 82260: gap of unknown length
            * 82261 167553: contig of 85293 bp in length.
            *
            * Location/Qualifiers
            1..167553
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /chromosome="5"
            /clone="CTD-2244D9"
            /clone_lib="CalTech human BAC library D"
            /clone_lib="50416 t 261 others"
            BASE COUNT 48962 a 35722 c 32192 g 50416 t 261 others
            ORIGIN
            Query Match 64.1%; Score 641.4; DB 2; Length 167553;
            Best Local Similarity 89.3%; Pred. No. 1.8e-190;
            Matches 737; Conservative 0; Mismatches 81; Indels 7; Gaps 4;
            Qy 183 AAAAGTCACGGAGCCTCAGTGTCCCTTCTGTGGAGCCAGAACCTCATGACGTCTAAGT 242
            Db 61428 AAAGTCACTAGACCTCAGTGTCCCTTCTGTGGAGCCAGAACCTCATGACGTCTAAGT 61369
            Qy 243 CTTGTTGTATGAACATGTCTCAGCCTGCGGCCCTGGTGGTGGTGCAGCATAGGAAGTA 302
            Db 61368 CTTGTTTATGACATGTCTCAGCCTGCGGCCCTGGTGGTGGTGCAGCATAGGAAGTA 61309
            Qy 303 TAAGGATGAGGTCTAGTTCATGGGCATGGAGCCTTCTCATTAATCTTGGCTGTCTGCC 362
            Db 61308 TAAGGATGAGGTCTAGTTCATGGGCATGGAGCCTTCTCATTAATCTTGGCTGTCTGCC 61249
            Qy 363 TTCTAGGGA-ATATAATCAACACTAATAAAGGAGGAGGTGAGAGCTGGCGCTGTGCGT 421
            Db 61248 TCTTAAGAACATATAATCAACACTAATAAAGGAGGAGGTGAGAGCTGGCGCTGTGCGT 61189
            Qy 422 TTGAGGAGGATGGCATGTGAAGTCACTGACCCCGTGGGAGGAGACACTCCCTGGCTC 481
            Db 61188 TTGAGGAGGATGGCATGTGAAGTCACTGACCCCGTGGGAGGAGACACTCCCTGGCTC 61129
            Qy 482 CATCTCTGCATCTTAGATTTATTTGGGACAGTTTATGATACACAGAGAGGAGAGCCCAT 541
            Db 61128 CATCTCTGCATCTTAGATTTATTTGGGAGGTTTATGATACACAGAGAGGAGAGCCCAT 61069
            Qy 542 CCATGAGGAGGTTGATTTAGATGAATATAATCAATGATAATTTCTTAGAGGAGGACTT 601
```

```
Db 61068 CCCAATGAGGGTTTGGATTAGGGGAATAGAAATCAATGATAAATCCTTAGAGGAGGACTG 61009
Qy 602 TTTATAATCAATC-TGAGAACAGGTTGAGCTACATGGATTGAGGGAGGGTGGAGC 660
Db 61008 TTTATATCAACTCTTGAGAACAGGTTGGGCTACATGGATTGAGGGAGGGTAGGAC 60949
Qy 661 CCCTTTAAAGAAAGCCCCAGAGACTGCCCTGCGCTCTCTCTCCGCCACAAGTTCCATT 720
Db 60948 CCCTTTAAAGAAAGCCCTAGAAATGTCCCTACCCCTTACACCCCCACAAAGTTCCCTT 60889
Qy 721 TATATATCTCCACCAGAGCTGTGAGAAATCCTGCGCTTCC-----GTCTCCAGATCAAG 776
Db 60888 TGTCAATCTCCACCAGAGCTGTGAGAAATCCTGCGCTTCCCTCTCTGCTCCAGATCAAG 60829
Qy 777 TCCTTCAGAAATGCAACTTACTTTCAGTCAGAGAGATATATCATCTTCTGACAGAGA 836
Db 60828 TCCCTCAGAAATGAGAGCTGTCTAGTCAGACAGAGATATCGTCACTCTGACTGAGGA 60769
Qy 837 GGAATTTGGGTTTGGTCCCGAGTCCATGAAGTGGCAGACTCAGATAAAGGTGAGAGCT 896
Db 60768 GGAATTTGGGTTTGGTCCAGTGCATGAAGTGCACAGTCAGATAAAGATGAGGCC 60709
Qy 897 TAGGAGATTAGGGAGGTGAGAACACTCTGTCTTGTGACCACTTCAGAGCCCTGG 956
Db 60708 TAGCAGATTAGCAAGACTAGGAGAAGACTCTATCTTGTGCGCAGCTTCAGAGAACCCTGG 60649
Qy 957 GGCCATGCTTCCCTGGTCAACATTAAGGCCCTGCTGCATGCTGACC 1001
Db 60648 GGCCATGCTTCCCTGGTCAACATTAAG--CCTGCTGCATGCGGAGC 60605

RESULT 9
AC002042 185994 bp DNA linear HTG 23-NOV-1999
LOCUS Homo sapiens clone A-180G2, *** SEQUENCING IN PROGRESS ***, 5
DEFINITION Homo sapiens clone A-180G2, *** SEQUENCING IN PROGRESS ***, 5
unordered pieces.
AC002042
VERSION AC002042.1 GI:3075381
KEYWORDS HTG: HTGS, PHASE1.
SOURCE Homo sapiens.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
1 (bases 1 to 185994)
Lofthus, B.J., Kim, U.J., Sneddon, V.P., Kalush, F., Brandon, R.,
Fuhrmann, J., Mason, T., Crosby, M.L., Barnstead, M., Cronin, L.,
Deslattes Mays, A., Cao, Y., Xu, R.X., Kang, H.L., Mitchell, S.,
Eichler, E., Harris, P.C., Venter, J.C. and Adams, M.D.
Genome duplications and other features in 12 Mb of DNA sequence
from human chromosome 16p and 16q
Genomics 60 (3), 295-308 (1999)
99425270
MEDLINE
PUBMED 10493829
2 (bases 1 to 185994)
Adams, M.D., Loftus, B.J., Zhou, L., Phillips, C., Brandon, R.C.,
Fuhrmann, J., Kim, U.J., Kerlavage, A.R. and Venter, J.C.
Human chromosome 16p13 BAC clone C17987SK-A-180G2
Unpublished
3 (bases 1 to 185994)
Adams, M.D. and Loftus, B.J.
Direct Submission
Submitted (29-APR-1997) The Institute for Genomic Research, 9712
Medical Center Dr., Rockville, MD 20850, USA
On Apr 22, 1998 this sequence version replaced gi:3068565.
* NOTE: This is a 'working draft' sequence. It currently
* consists of 5 contigs. The true order of the pieces
* is not known and their order in this sequence record is
* arbitrary. Gaps between the contigs are represented as
* runs of N, but the exact sizes of the gaps are unknown.
* This record will be updated with the finished sequence
* as soon as it is available and the accession number will
* be preserved.
```

```
* 1 2198: contig of 2198 bp in length
* 2199 2248: gap of unknown length
* 2249 46288: contig of 44040 bp in length
* 46289 46338: gap of unknown length
* 46339 110919: contig of 64581 bp in length
* 110920 110969: gap of unknown length
* 110970 145225: contig of 34256 bp in length
* 145226 145275: gap of unknown length
* 145276 185994: contig of 40719 bp in length.

FEATURES
    Location/Qualifiers
        source
            1..185994
            /organism="Homo sapiens"
            /db_xref="taxon:9606"
            /clone="A-180G2"

BASE COUNT 51275 a 43334 c 43047 g 48137 t 201 others

ORIGIN
Query Match 64.1%; Score 641.4; DB 2; Length 185994;
Best Local Similarity 89.3%; Pred. No. 1.8e-190;
Matches 737; Conservative 0; Mismatches 81; Indels 7; Gaps 4;

Qy 183 AAAAGTCACGGAGCCTCAGTGTCCCTCTGTGGAGCCCGACACCTGATGACAGTCTAAGT 242
Db 26298 AAAGTCTACTGAGCCTCAGTGTCCCTCTGTGGAGCCCGACACCTGATGACAGTCTAAGT 26357
Qy 243 CTTGTGTATGAACATGCTTGCACCTGCGGCCCTGTGTGGTGTGTCAGCATAGGAAGTA 302
Db 26358 CTTGTGTATGACATGCTTGCACCTGCGGCCCTGTGTGGTGTGTCAGCATAGGAAGTA 26417
Qy 303 TAAGGATGAGGTCTAGTGTGCGCCATGGAGCCCTTCTCAATTAATCTTGGCTGTCTGCC 362
Db 26418 TAAGGATGAGGTCTAGTGTGCGCCATGGAGCCCTTCTCAATTAATCTTGGCTGTCTACC 26477
Qy 363 TTCTAGGGA-ATATAATCAACACTAATAAGAGAGAAAGTGTGACAGCTGGCCCTGTCTCT 421
Db 26478 TCCTAAGAACATATAATCAACACTAATAAGAGAAAGTGTGACAGCTGGCCCTGTCTCTCT 26537
Qy 422 TTGAGGAGGATGGCGATGTGAAAGTGTGACACCTGTGGGAGGAGACACTCCCTGGGCTC 481
Db 26538 TTGAGGAGGATGGCGATGTGAAAGTGTGACACCTGTGGGAGGAGACACTCCCTGGGCTC 26597
Qy 482 CATCTCTGCATCTTAGATTATTTGGGACAGTTTGTATACACAGAGAAGAGAGACCAT 541
Db 26598 CATCTCTGCATCTTAGATTATTTGGGAGGTTTGTATACACAGAGAAGAGAGACCAT 26657
Qy 542 CCCATGAGGGTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 601
Db 26658 CCCATGAGGGTTTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGATG 26717
Qy 602 TTTATAATCAATC-TGAGAACAGGTTGAGCTACATGGGATTGGAGGGAGGGTGGAGC 660
Db 26718 TTTATAATCAATC-TGAGAACAGGTTGAGCTACATGGGATTGGAGGGAGGGTGGAGC 26777
Qy 661 CCCTTTAAAGAAAGCCCGAGAGACTGCCCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 720
Db 26778 CCCTTTAAAGAAAGCCCGAGAGACTGCCCTGCGCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT 26837
Qy 721 TATATATCTCCACCAGAGCTGTGAGAAATCCTGCGCTTCC-----GTCTCCAGATCAAG 776
Db 26838 TGTCAATCTCCACCAGAGCTGTGAGAAATCCTGCGCTTCC-----GTCTCCAGATCAAG 26897
Qy 777 TCCTTCAGAAATGCAACTTACTTTCAGTCAGAGAGATATATCATCTTCTGACAGAGA 836
Db 26898 TCCTTCAGAAATGCAACTTACTTTCAGTCAGAGAGATATATCATCTTCTGACAGAGA 26957
Qy 837 GGAATTTGGGTTTGGTCCCGAGTCCATGAAGTGGCAGACTCAGATAAAGGTGAGAGCT 896
Db 26958 GGAATTTGGGTTTGGTCCAGTGCATGAAGTGCACAGTCAGATAAAGATGAGGCC 27017
Qy 897 TAGGAGATTAGGGAGGTGAGAACACTCTGTCTTGTGACCACTTCAGAGAGCCCTGG 956
Db 27018 TAGCAGATTAGCAAGACTAGGAGAAGACTCTATCTTGTGCGCAGCTTCAGAGAACCCTGG 27077
```



REFERENCE 4 (bases 1 to 144577)  
 AUTHORS DOE Joint Genome Institute.  
 TITLE Direct Submission  
 JOURNAL Submitted (06-AUG-2002) Production Sequencing Facility, DOE Joint Genome Institute, 2800 Mitchell Drive, Walnut Creek, CA 94598, USA  
 COMMENT On Aug 6, 2002 this sequence version replaced gi:21397290.  
 Sequence Quality Assessment:  
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.  
 -----  
 Sequence Quality Assessment:  
 This entry has been annotated with sequence quality estimates computed by the Phrap assembly program. All manually edited bases have been reduced to quality zero. Quality levels above 40 are expected to have less than 1 error in 10,000 bp.  
 Base-by-base quality values are not generally visible from the GenBank flat file format but are available as part of this entry's ASN.1 file.  
 -----  
 FEATURES  
 source Location/Qualifiers  
 1..144577  
 /organism="Homo sapiens"  
 /db\_xref="taxon:9606"  
 /chromosome="16"  
 /clone="RP11-244B22"  
 BASE COUNT 39644 a 31376 c 31688 g 41869 t  
 ORIGIN  
 Query Match 56.9%; Score 569.6; DB 9; Length 144577;  
 Best Local Similarity 84.5%; Pred. No. 8e-168;  
 Matches 700; Conservative 0; Mismatches 119; Indels 9; Gaps 5;  
 QY 183 AAAAGTCACGGAGCTCAGTCTCCCTCTCTGTGGAGCCAGAACCTGATCGAGGCTCAAGT 242  
 Db 65205 AAAAGTCACGGAGCTCAGTCTCCCTCTCTGTGGAGCCAGAACCTGATCGAGGCTCAAGT 65146  
 QY 243 CCTGTTGATGAACATGCTCAGTCCCTCTCTGTGGAGCCAGAACCTGATCGAGGCTCAAGT 302  
 Db 65145 CCTGTTGATGCTATGCGGTGACCTGCTGGCCCTGCTGGTGTGGAGCATAGGAAGTA 65086  
 QY 303 TAAGGATGAGGCTGATGATGCGCATGAGGACCTTTCTCTAATCTTGGCTGTCTGCC 362  
 Db 65085 CATGGGATGAGGCTGATGATGCGCATGAGGACCTTTCTGAGGATCTTGGCTGTGCC 65026  
 QY 363 TTCTAGG-GAATATAATCAACACTAA-TAAAGGAGGAGGTGAGCAGCTGCGCTGTCCG 420  
 Db 65025 TTCTAGGAAATATAATCAACACTAAAGTAAAGGAGTGAATGAGCCTTGGGGCCCTCAC 64966  
 QY 421 TTTGAGGAGGATGCGGATGTAAGTCAGTGACACCGTGGGAGGACACTCCCTGCTGCT 480  
 Db 64965 TGTGAGGAGGCTGGGATGTAAGTCAGTGACACCGTGGGAGGACACTACCTGCTGCT 64906  
 QY 481 CCATCCTCTGCTATGATTTATTTGGACAGCTTTGATACACAGAGAGGAGACCCA 540  
 Db 64905 CCATCCTCTGCTATGATTTGTTGGAGGTTTGTGTTTATAGAGAGGAGGAGACCCA 64846  
 QY 541 TCCCAATGAGGGTTGATTAAGATGAATATAATCAATCAATTAATTCCTAGAGGAGGACT 600  
 Db 64845 TTGAGTGGAGGGTTGATTAAGGAACTAGAACTAAATTAAGTCTTATAGGAGGACT 64786  
 QY 601 TTTTATAATCAACTCTGAGACAGCTGGAGCTACATGGGATTTGGAGGGAGGCTGGAGC 660  
 Db 64785 GTTTATATCAACTCTGAGACAGCTGGTGGCACTACATGGGATTTGGAGGGATGATGAAC 64726  
 QY 661 CCCTTAAGAAAGACCCCAAGAGACT-GCCCTCCCTCTCTCTCCCTCCCAAGTTCCAT 719  
 ||||||| || |||||||||||||||| || |||||| || |||||||||||| |||

Db 64725 CCCTTAAGGGATAAGCCCCAGAGAGACTGGCTTCTGCTCCTCCCTCCCAACAATTCCT 64666  
 QY 720 TTATTATCTTCCACCCAGGAGCTGTGAGAATCCTGCCCT----TCCGCTCTCAGATCAA 775  
 ||||| |||||||||||||||| |||||||||||| || |||||| ||||||||  
 Db 64665 TTATCATCTTCCACCCAGGAGCTGTGAGAATCCTGCTCTCTGCTGTCTGTAGATCAA 64606  
 QY 776 GTCCTTTCAGGAATGCAACTACTTTCAGTGACAAGAGATAAATATCATCTTCTGCAGAGG 835  
 ||||| |||||||| |||||||||||| |||||||||||| |||||||| |||||  
 Db 64605 ATCTCTCAAGAATGCACTGCTTTCAGTGACAAGAGATAAATGTCATCTTCCAAGTGA 64546  
 QY 836 AGGAATTTGGGGTTGGTCCAGTCCATGAA--GTGGCACAGTCAGATAAAGGTGAGA 893  
 ||||| |||||||| |||||||||||| |||||||||||| |||||||| |||||  
 Db 64545 AAGAATTTGGGGTTGGTTCAGATTCATGAAGTGTGACACAGTCAGATAAAGGTGAGA 64486  
 QY 894 GCTTAGGAGATTAGCGGAGGTGAGAAGAACACTGCTCTTGTGACCACTTCAGAGAGCC 953  
 ||||| |||||||| |||||||||||| |||||||||||| |||||||| |||||  
 Db 64485 CGGTAGCAGATTAGCAGAGGTGAGGAAGACTCCATCTTGGGCCAGCTTCAGAAAGCC 64426  
 QY 954 TGGGCCCATGCTTCTCTGTCACCATTAAGGCCCTGCTCATGTTGATGACC 1001  
 ||||| |||||||| |||||||||||| |||||||||||| |||||||| |||||  
 Db 64425 TGTGCTATGCTCCCTGTCACCATTTGGCCCTGTTGTCATGGGGACC 64378  
 RESULT 12  
 AL390198  
 LOCUS AL390198 167717 bp DNA linear PRI 04-APR-2001  
 DEFINITION Human DNA sequence from clone RP13-401N8 on chromosome 20 Contains three putative novel genes, a pheromone receptor pseudogene (PHRET), a novel zinc-finger pseudogene, nine CpG islands, ESTs, STSS and GSSs, complete sequence.  
 ACCESSION AL390198  
 VERSION AL390198.17 GI:11967552  
 KEYWORDS HTG; CpG island; pheromone receptor; PHRET; zinc-finger.  
 SOURCE human.  
 ORGANISM Homo sapiens  
 Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.  
 1 (bases 1 to 167717)  
 Bird.C.  
 Direct Submission  
 Submitted (21-MAR-2001) Sanger Centre, Hinxton, Cambridgeshire, CB10 1SA, UK. E-mail enquiries: humquerry@sanger.ac.uk  
 requests: clonerequest@sanger.ac.uk  
 On Dec 22, 2000 this sequence version replaced gi:11557932.  
 During sequence assembly data is compared from overlapping clones. Where differences are found these are annotated as variations together with a note of the overlapping clone name. Note that the variation annotation may not be found in the sequence submission corresponding to the overlapping clone, as we submit sequences with only a small overlap as described above.  
 The following abbreviations are used to associate primary accession numbers given in the feature table with their source databases: EMBL, EMBL; Sw., SWISSPROT; Tr., TREMBL; Wp., WORMPEP; Information on the WORMPEP database can be found at <http://www.sanger.ac.uk/projects/C-elegans/wormpep> This sequence was generated from part of bacterial clone contigs of human chromosome 20, constructed by the Sanger Centre Chromosome 20 Mapping Group. Further information can be found at <http://www.sanger.ac.uk/HGP/Chr20>  
 RP13-401N8 is from the library RP13-13.2 constructed by the group of Pieter de Jong. For further details see <http://www.chori.org/bacpac/home.htm>  
 VECTOR: pBACE3.6  
 IMPORTANT: This sequence is not the entire insert of clone RP13-401N8 it may be shorter because we sequence overlapping sections only once, except for a 100 base overlap.  
 The true left end of clone RP4-760C5 is at 167618 in this sequence. The true right end of clone RP4-694B14 is at 100 in this sequence. This sequence was finished as follows unless otherwise noted: all regions were either double-stranded or sequenced with an alternate chemistry or covered by high quality data (i.e., phred quality >= 30); an attempt was made to resolve all sequencing problems, such as compressions and repeats; all regions were covered by at least









Tateno, M., Catanese, J.J. and de Jong, P.J. (1998) An improved approach for construction of bacterial artificial chromosome libraries. Genomics 51:1-8. The clone may be obtained either from Research Genetics, Inc. (<http://www.resgen.com>) or Pieter de Jong and coworkers at the Roswell Park Cancer Institute (<http://bacpac.med.buffalo.edu>)

VECTOR: pBACE3.6

#### NEIGHBORING SEQUENCE INFORMATION:

Actual start of this clone is at base position 1 of RP11-114G11; actual end is at base position 178105 of RP11-114G11.

The sequence RP11-114G11 contains a dinucleotide (AT) repeat from base position 161700 to 163400 where there are low quality and single chemistry regions. Assembly is consistent with PCR and digest information.

#### FEATURES

Location/Qualifiers

Source

```
1..178105
/organism="Homo sapiens"
/db_xref="taxon:9606"
/chromosome="7"
/map="7"
/clone="RP11-114G11"
/clone_lib="RPCI-11"

misc_feature
1..801
/note="similar to EST AU132789 (NID:g10993328)"

misc_feature
1..665
/note="similar to EST AU123664 (NID:g10948380)"

misc_feature
1..623
/note="similar to EST AU143734 (NID:g11005255)"

misc_feature
1..589
/note="similar to EST BE870310 (NID:g10319086)"

misc_feature
1..577
/note="similar to EST BE897039 (NID:g10362079)"

misc_feature
1..440
/note="similar to EST AI241755 (NID:g3837152) qu71f07.x1"

misc_feature
2..663
/note="similar to EST AU123448 (NID:g10948164)"

misc_feature
2..631
/note="similar to EST AI566486 (NID:g4524938) tr60d09.x2"

misc_feature
2..629
/note="similar to EST AI733621 (NID:g5054734) an30d02.x5"

misc_feature
2..535
/note="similar to EST AA205471 (NID:g1803461) zq66a11.r1"

misc_feature
2..510
/note="similar to EST AW835596 (NID:g7929570)"

misc_feature
3..527
/note="similar to EST BE064551 (NID:g8409201)"

misc_feature
3..522
/note="similar to EST AW291262 (NID:g6697898)"

misc_feature
3..510
/note="similar to EST AI732980 (NID:g5054093) og79e08.x5"

misc_feature
3..510
/note="similar to EST AI788680 (NID:g5336396) uk52f01.x1"

misc_feature
6..490
/note="similar to EST AA397793 (NID:g2051001) zt88e05.r1"

misc_feature
8..600
/note="similar to EST AV755078 (NID:g10912926)"

misc_feature
12..630
/note="similar to EST AL042967 (NID:g5422398)"

misc_feature
12..630
/note="similar to EST BE396253 (NID:g9341618)"

misc_feature
12..629
/note="similar to EST BF240985 (NID:g11154910)"

misc_feature
17..577
/note="similar to EST AA976492 (NID:g3152284) oq71e09.s1"

misc_feature
18..461
/note="similar to EST AA743478 (NID:g2782984) ny21g03.s1"

misc_feature
28..732
/note="similar to EST AA768909 (NID:g2820147) nz82h01.s1"

misc_feature
29..622
/note="similar to EST AI476267 (NID:g4329312) ti72g01.x1"

misc_feature
31..467
/note="similar to EST N98899 (NID:g1270324) za68f06.r1"
```

```
misc_feature
32..567
/note="similar to EST AW392975 (NID:g6897634)"

misc_feature
33..357
/note="similar to EST AI042041 (NID:g3281235) ox60b08.x1"

misc_feature
36..491
/note="similar to EST AI744718 (NID:g5112928) wf90c08.x1"

misc_feature
41..675
/note="similar to EST AV704018 (NID:g10721339)"

misc_feature
50..537
/note="similar to EST AW859939 (NID:g7955632)"

misc_feature
62..479
/note="similar to EST BF146315 (NID:g11027710) uy27d08.x1"

misc_feature
62..442
/note="similar to EST T47251 (NID:g649233) yb64d02.s1"

misc_feature
68..720
/note="similar to EST BF020339 (NID:g10751671) uw64e04.x1"

misc_feature
71..704
/note="similar to EST AI680077 (NID:g4890259) tw64g09.x1"

misc_feature
79..458
/note="similar to EST T05813 (NID:g316963)"

misc_feature
86..701
/note="similar to EST AI554749 (NID:g4487112) tn27h04.x1"

misc_feature
86..522
/note="similar to EST AI049550 (NID:g3298667) an30d02.x1"

misc_feature
86..466
/note="similar to EST BE878800 (NID:g10327576)"

misc_feature
92..408
/note="similar to EST BF247382 (NID:g11162748)"

misc_feature
121..440
/note="similar to EST AI081664 (NID:g3418456) ou63e05.s1"

misc_feature
122..472
/note="similar to EST BE082082 (NID:g8472378)"

misc_feature
125..440
/note="similar to EST AI758269 (NID:g5151992) ty05c02.x1"

misc_feature
126..435
/note="similar to EST AI049825 (NID:g3298942) an37f07.x1"

misc_feature
141..440
/note="similar to EST AI281694 (NID:g3919927) qt67e11.x1"

misc_feature
158..787
/note="similar to EST AW272732 (NID:g6659689) xu22a04.x1"

misc_feature
228..522
/note="similar to EST AW887346 (NID:g8049359)"

misc_feature
571..798
/note="similar to EST BF247382 (NID:g11162748)"

misc_feature
571..724
/note="similar to EST AI744718 (NID:g5112928) wf90c08.x1"

misc_feature
854..991
/note="similar to EST BF247382 (NID:g11162748)"

misc_feature
1836..2342
/note="similar to EST BF002402 (NID:g10702677)"

repeat_region
2702..2795
/rpc_family="L2"

repeat_region
3097..3210

Query Match      33.0%; Score 330.6; DB 9; Length 178105;
Best Local Similarity 71.8%; Pred. No. 2.3e-92;
Matches 587; Conservative 0; Mismatches 184; Indels 46; Gaps 10;

QY  206  CCTTCTGGAGCCAGACCTGATGCAGGCTCTAAGTCTGTGTGTATGAACATGCTCTGA 265
      ||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 125515  CCTGCTAGAGCCAGACGCTGATGCAGGTCCTCTCTATGCACATATTTTCAC 125574

QY  266  CCCTGGCGCCCTGCTGGTGGTCAGCATAGGAAGTATAGGTCAGTCTAGTCATGG 325
      ||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 125575  CCCTGCTGGCTG-----GCAGTGTGGAGGCTGGGGGATGGGCAAGTCATGG 125625

QY  326  GCGATGAGGACCTTCTCAATTAATCTTGCTGCTCCCTCTTAGGG-AAATAATCAACAC 384
      ||| | ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 125626  GCCA-GGGTCTCTTCTGAGGCTCTTGGCTGGCTCTCTCCAGGGAATGTGATCAACAC 125684

QY  385  TAATAAAGAGGAGGTCAGCAGCTGGCGCTGTCGCTTTGAGGGAGGATGGCGATGTGAA 444
```

```
Db 125685 CCAGAAAGGAAAGGCGCAGCTGGGCGCCCTCCTCTTTGGGGAGGATGGGATGTGGA 125744
Qy 445 AGTCAGTGACC-----ACGCTGGGAGGACACTCCTCGCTCCATCCTCTGCAT 493
Db 125745 AGTCAGAGCCCCACCACCTGAGTGGGTGAGAACATCCCCGGCTCCATCCTCTGCAT 125804
Qy 494 CTTAGATTATTGGGACAGTTTGATACAGAGAGGAGAGACCCATCCCAATGGAGGG 553
Db 125805 CTTGATTATTGGGAGGTTTGATACAG---AGGACGAGACCCATCCCAATGGAGGG 125861
Qy 554 TTTGATTAGATCAATATATATCAATGATAAATCTTAGAGAGG-----GACTTTTATA 607
Db 125862 TGTGGGTAGGAGACAGGAATCTCGACAAATCTTAGCGGGGCTGTTTCTCTGTTTATA 125921
Qy 608 ATCAACTCTGAGAACAGCTGGAGCTACATGGATTGAGGAGGGGAGGAGCCCTTAA 667
Db 125922 TCCAACTCTGAGAAAGGCTGGGCTGCTAGGATTGAGAGAGGGGA---CCTACTAA 125978
Qy 668 AAGAAAGCCCGAGAGCTGCCCTCGCTCTCTCTCCCCACAAAGTTTCCATTTATTATC 727
Db 125979 GAGAAAGCCCTAGATGATCAGCCC--CAGTTTACTCCCTCCCAAGGCCCTTCGGCATC 126036
Qy 728 TTCCACCAGAGCTGTCAGAAATCCTG-----CCCTTCGCTCCAGATCAAAATGCC 779
Db 126037 TTCCACCAGAGCCCTGTCAGGCTCTCTGTTTCCCTCTGCTCCAGAGATTAAGTCC 126096
Qy 780 TTCAGGAATCAACTCTCAGTCAGACAGATAATATCATCTCTCAGACAGAGGGA 839
Db 126097 TCCAGAAATGAGCTGCTCCGTGACAGAGGGAATATCATCTTCGACTCAGAGG- 126155
Qy 840 ATTTGGGTTTGTCCAGTCCATCAAGTGGCAGACAGTACAGATAAAAGGTGAGACTTAG 899
Db 126156 -TATGGGTTGTTTCAAGTCCGTGGAGCCGACACATGATTGAGAGGTGAGACCTAG 126214
Qy 900 GAGATTAGCGGAGTAGAAGAACACTCTGCTGTGTACACAGCTTTCAGAGAGCCCTGGGGC 959
Db 126215 CGGATTGACAGAGGTTGGGCGGAGACTCTGCTGTTGTGTCAGTTTCAGACATCCTGTGGC 126274
Qy 960 CATGCTCTCCTGTTCAACATTAGGCCCTGCTGCATCG 996
Db 126275 CACGGCTCCCTGTCAGCAGCAGGCCCCCTGTTGAATGG 126311

RESULT 14
AL355134/c
LOCUS AL355134 140653 bp DNA linear HTG 12-SEP-2001
DEFINITION Homo sapiens chromosome 13 clone RP11-398A7, *** SEQUENCING IN
PROGRESS ***, 18 unordered pieces.
ACCESSION AL355134
VERSION AL355134.11 GI:15617248
KEYWORDS HTG; HTGS_PHASE1; HTGS_CANCELLED.
SOURCE human.
ORGANISM Homo sapiens
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.
Pearce, A.
Direct Submission
Submitted (11-SEP-2001) Sanger Centre, Hinxton, Cambridgeshire,
CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk
requests: clonerequest@sanger.ac.uk
On Sep 13, 2001 this sequence version replaced gi:13477090.
----- Genome Center
Center: Sanger Centre
Center code: SC
Web site: http://www.sanger.ac.uk
Contact: humquery@sanger.ac.uk
----- Project Information
Center project name: BA398A7
----- Summary Statistics
Assembly program: XGAP4; version 4.5
Sequencing vector: plasmid; L08752; 100% of reads
```

Chemistry: Dye-terminator Big Dye; 98% of reads  
Chemistry: Dye-primer Big Dye; 1% of reads  
Consensus quality: 135106 bases at least Q40  
Consensus quality: 136089 bases at least Q30  
Consensus quality: 137104 bases at least Q20  
Insert size: 138953; sum-of-contigs  
Insert size: 180323; 4.8% error; agarose-fp  
Quality coverage: 6.90x in Q20 bases; sum-of-contigs Quality  
coverage: 6.28x in Q20 bases; agarose-fp  
-----

\* NOTE: This is a 'working draft' sequence. It currently  
\* consists of 18 contigs. The true order of the pieces  
\* is not known and their order in this sequence record is  
\* arbitrary. Gaps between the contigs are represented as  
\* runs of N, but the exact sizes of the gaps are unknown.  
\* This record will be updated with the finished sequence  
\* as soon as it is available and the accession number will  
\* be preserved.

1 14926: contig of 14926 bp in length  
\* 14927 15026: gap of 100 bp  
\* 15027 29840: contig of 14814 bp in length  
\* 29841 29940: gap of 100 bp  
\* 29941 38295: contig of 8355 bp in length  
\* 38296 38395: gap of 100 bp  
\* 38396 40640: contig of 2245 bp in length  
\* 40641 40740: gap of 100 bp  
\* 40741 66942: contig of 26202 bp in length  
\* 66943 67042: gap of 100 bp  
\* 67043 74245: contig of 7203 bp in length  
\* 74246 74345: gap of 100 bp  
\* 74346 77014: contig of 2669 bp in length  
\* 77015 77114: gap of 100 bp  
\* 77115 79339: contig of 2225 bp in length  
\* 79340 79439: gap of 100 bp  
\* 79440 81637: contig of 2198 bp in length  
\* 81638 81737: gap of 100 bp  
\* 81738 84332: contig of 2595 bp in length  
\* 84333 84432: gap of 100 bp  
\* 84433 91475: contig of 7043 bp in length  
\* 91476 91575: gap of 100 bp  
\* 91576 94846: contig of 3271 bp in length  
\* 94847 94946: gap of 100 bp  
\* 94947 110268: contig of 15322 bp in length  
\* 110269 110368: gap of 100 bp  
\* 110369 112370: contig of 2002 bp in length  
\* 112371 112470: gap of 100 bp  
\* 112471 114853: contig of 2383 bp in length  
\* 114854 114953: gap of 100 bp  
\* 114954 118294: contig of 3341 bp in length  
\* 118295 118394: gap of 100 bp  
\* 118395 137994: contig of 19600 bp in length  
\* 137995 138094: gap of 100 bp  
\* 138095 140653: contig of 2559 bp in length.

## FEATURES

source  
1. .140653  
/organism="Homo sapiens"  
/db\_xref="taxon:9606"  
/chromosome="13"  
/clone="RP11-398A7"  
/clone\_lib="RPCI-11.2"  
1. .14926  
/note="assembly\_fragment:00489  
fragment\_chain:1"  
15027. .29840  
/note="assembly\_fragment:00109  
fragment\_chain:1"  
29941. .38295  
/note="assembly\_fragment:00395  
fragment\_chain:1"  
38396. .40640  
/note="assembly\_fragment:01907  
fragment\_chain:2"  
40741. .66942  
misc\_feature  
misc\_feature  
misc\_feature  
misc\_feature  
misc\_feature

|   |        |  |        |
|---|--------|--|--------|
| QY  | 708    | CACAAAGTTCCATTTATTATCTTCCACCAGGAGCTGTGAGAACTCCTCCCTTCCGCTCT--  | 765    |
| Db  | 133605 | TATAAGGCCCTTACAGAGTCTTTTCCACCAGCCCTGTGTCAGATCCTGTCTTCCCTCTGT   | 133546 |
| QY  | 766    | -----CCAGATCAAAGTCTCTTCAGGAAATGCAACTACTTTCAGTGA-CAAGAGATAATTAT | 819    |
| Db  | 133545 | TTCCAGAAGATTAAAGTCTCTCCAGGAGATGCAGCAGTTCCACACAGCTTGGAAACCATCAT | 133486 |
| QY  | 820    | CATCTTCTGACAGAGAGGAATTTGGGGTTTGGTCCCACTCCATGAAGTGGCAGAGTCAG    | 879    |
| Db  | 133485 | CATCTTCAGACTCAAGAGGAATTTTCGGGCTTTGTTCCAAGCC-TGGAGCAGCAATCAG    | 133427 |
| QY  | 880    | AATAAAGGTGAGAGCTTAGGAGATTAGCGGAGGGTGAAGAACAACTCTGTCTTTGTGACC   | 939    |
| Db  | 133426 | AATAAAGGCAAGACCTTAGCAGATGAGCAGAGGGTAGGAGGGGAGACTGCTTCTGCCGCC   | 133367 |
| QY  | 940    | AGCTTCAGAGAGCCTGGGGCCATGGCTTCCTGGTCAACATTAGGCCCTGCTGCATGGTGA   | 999    |
| Db  | 133366 | AGCCTCACACAGCGTGTGGCCATGGTTCCTTCGGCCGATCAGGTCCTGTTGCACCTGGA    | 133307 |
| QY  | 1000   | C 1000   |        |
| Db  | 133306 | C 133306   |        |
| RESULT 15   |        |  |        |
| AL353644  |        |  |        |
| LOCUS   |        |  |        |
| DEFINITION  |        |  |        |
| Homo sapiens chromosome 13 clone RP11-164K15, *** SEQUENCING IN   |        |  |        |
| PROGRESS ***, 13 unordered pieces.                                |        |  |        |
| AL353644  |        |  |        |
| ACCESSION   |        |  |        |
| VERSION   |        |  |        |
| AL353644.4 GI:9926526   |        |  |        |
| KEYWORDS  |        |  |        |
| HTG; HTGS_PHASE1; HTGS_DRAFT.                                     |        |  |        |
| SOURCE  |        |  |        |
| human.  |        |  |        |
| - ORGANISM  |        |  |        |
| Homo sapiens  |        |  |        |
| Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; |        |  |        |
| Mammalia; Eutheria; Primates; Catarrhini; Hominidae; Homo.        |        |  |        |
| REFERENCE   |        |  |        |
| 1 (bases 1 to 150286)   |        |  |        |
| AUTHORS   |        |  |        |
| Burton, J.  |        |  |        |
| TITLE   |        |  |        |
| Direct Submission   |        |  |        |
| JOURNAL   |        |  |        |
| Submitted (20-JAN-2001) Sanger Centre, Hinxton, Cambridgeshire,   |        |  |        |
| CB10 1SA, UK. E-mail enquiries: humquery@sanger.ac.uk             |        |  |        |
| requests: clonerequest@sanger.ac.uk                               |        |  |        |
| On Aug 25, 2000 this sequence version replaced gi:92131118.       |        |  |        |
| COMMENT   |        |  |        |
| ----- Genome Center   |        |  |        |
| Center: Sanger Centre   |        |  |        |
| Center code: SC   |        |  |        |
| Web site: http://www.sanger.ac.uk                                 |        |  |        |
| Contact: humquery@sanger.ac.uk                                    |        |  |        |
| ----- Project Information   |        |  |        |
| Center project name: BA164K15                                     |        |  |        |
| ----- Summary Statistics  |        |  |        |
| Assembly program: XGAP4; version 4.5                              |        |  |        |
| Sequencing vector: plasmid; L08752; 100% of reads                 |        |  |        |
| Chemistry: Dye-terminator Big Dye; 100% of reads                  |        |  |        |
| Consensus quality: 143155 bases at least Q40                      |        |  |        |
| Consensus quality: 146193 bases at least Q30                      |        |  |        |
| Consensus quality: 147821 bases at least Q20                      |        |  |        |
| Insert size: 149086; sum-of-contigs                               |        |  |        |
| Insert size: 177098; 0.8% error; agarose-fp                       |        |  |        |
| Quality coverage: 3.91x in Q20 bases; sum-of-contigs Quality      |        |  |        |
| coverage: 3.37x in Q20 bases; agarose-fp                          |        |  |        |
| -----   |        |  |        |
| * NOTE: This is a 'working draft' sequence. It currently          |        |  |        |
| * consists of 13 contigs. The true order of the pieces            |        |  |        |
| * is not known and their order in this sequence record is         |        |  |        |
| * arbitrary. Gaps between the contigs are represented as          |        |  |        |
| * runs of N, but the exact sizes of the gaps are unknown.         |        |  |        |
| * This record will be updated with the finished sequence          |        |  |        |
| * as soon as it is available and the accession number will        |        |  |        |
| * be preserved.   |        |  |        |
| * 7337: contig of 7337bp in length                                |        |  |        |
| * 1   |        |  |        |



**THIS PAGE BLANK (USPTO)**